European Journal of Public Health, Vol. 19, No. 3, 271–277

doi:10.1093/eurpub/ckn144 Advance Access published on 22 January 2009

Optimising lifestyle interventions: identification of health behaviour patterns by cluster analysis in a German 50+ survey

Sven Schneider[1], Christina Huy[1], Marc Schuessler[1], Katharina Diehl[1], Stefanie Schwarz[2]

Background: Many prevention and intervention measures are still targeting isolated behaviours such as tobacco use or physical inactivity. Cluster analysis enables the aggregation of single health behaviours in order to identify distinctive behaviour patterns. The purpose of this study was to group a sample of the over-50 population into clusters that exhibit specific health behaviour patterns regarding regular tobacco use, excessive alcohol consumption, unhealthy diet and physical inactivity. **Methods:** From the total population of the federal state of Baden-Wuerttemberg, Germany, 982 men and 1020 women aged 50–70 were randomly selected. Subjects were asked by trained interviewers in computer-assisted telephone interviews (CATI) about health behaviour and sociodemographic characteristics. Cluster analysis was conducted to identify distinct health behaviour patterns. Multinomial logistic regression was used to characterize clusters by specific social attributes. **Results:** Five homogeneous health behaviour clusters were identified: 'No Risk Behaviours' (25.3%), 'Physically Inactives' (21.1%), 'Fruit and Vegetable Avoiders' (18.2%), 'Smokers with Risk Behaviours' (12.7%) and 'Drinkers with Risk Behaviours' (22.7%). Whereas the first cluster is the ideal in terms of risk and prevention, the latter two groups include regular users of tobacco and excessive consumers of alcohol, who also engage in other risk behaviours like inactivity and maintaining an unhealthy diet. These two risk groups also exhibit specific sociodemographic attributes (male, living alone, social class affiliation). **Conclusion:** Unhealthy behaviours evidently occur in typical combinations. An awareness of this clustering enables prevention and intervention measures to be planned so that multiple behaviours can be modified simultaneously.

Keywords: cluster analysis, health behaviour, lifestyle, prevention, risk factors.

Introduction

About 70–80% of deaths in developed countries are lifestyle-related.[1,2] In recognition of this fact, behavioural risk factors have long been at the centre of social and preventive medicine, epidemiology and public health research.[2–4] Nevertheless, many prevention and intervention measures are still focused solely on modifying isolated behaviours. It should come as no surprise that such a selective approach meets with only modest success in many cases.[5,6]

One way out of the limited approach of selective interventions is to focus on more complex behavioural patterns rather than on isolated behaviours. In terms of planning comprehensive prevention programmes and interventions, it would therefore be useful to know the extent to which the most important behavioural risk factors (regular tobacco use, excessive alcohol consumption, an unhealthy diet and physical inactivity) aggregate in certain sectors of the population and whether typical risk groups can be identified on that basis. The cluster analysis method enables this kind of holistic analysis and facilitates the identification of intervention-relevant target groups. Although this complex method is in widespread use in sociology[7–9] and commercial market research[10,11] it is rare in epidemiological and public health studies.

Previously published cluster analyses on this topic mainly come from the US, Europe and South America. They either determine health behaviour in young people[12,13] or the general population, (i.e. in young adults of unspecified ages up to and including the elderly).[6,8,14–21] Moreover, most cluster analyses are limited to the correlation between two behavioural risk factors and do not consider multidimensional clusters.[12–14,16–20] Only two publications focus specifically on seniors and describe the identification of clusters of multiple behaviours concurrently.[6,21]

Against the background of demographic change and a still rising life expectancy, this paper looks at the over 50-year-old population. This study strengthens the hypothesis that there is a strong correlation in the over 50-year-old population between the four major behavioural risk factors of regular tobacco use, excessive alcohol consumption, unhealthy diet and physical inactivity. Furthermore, the study demonstrates that it is possible to cluster the participants into several definable risk groups.

Methods

Study participants

The basis for this article is data taken from the cross-sectional study 'Living an Active Life', a collaborative project by the German universities of Heidelberg, Stuttgart and Tuebingen. Inclusion criteria for subjects were defined as follows: (i) aged 50–70 years and (ii) residence in Baden-Wuerttemberg, a federal state in southern Germany. Of the 10 313 randomly selected individuals, 882 subjects were excluded according to the following criteria: unable to speak or understand German

1 Mannheim Institute of Public Health, Social and Preventive Medicine (MIPH), University of Heidelberg, Mannheim Medical Faculty, Mannheim, Germany
2 Faculty of Social Sciences, University of Mannheim, Mannheim, Germany
Correspondence: Sven Schneider, Mannheim Institute of Public Health, Social and Preventive Medicine (MIPH), Mannheim Medical Faculty, University of Heidelberg, Ludolf-Krehl-Straße 7-11, D-68167 Mannheim, Germany. tel: +49 621 383 99 17, fax: +49 621 383 99 20, e-mail: sven.schneider@medma.uni-heidelberg.de

($n = 637$), dead, severely ill or senile ($n = 137$), away for an extended period of time ($n = 52$) or hearing impaired ($n = 56$). A response rate of 21.23% was achieved in the study, which corresponded to a net sample size of 2002 persons, 982 men and 1020 women.

To ensure that the descriptive part of the study was representative, the dataset was weighted. The weighting factor was determined in cooperation with ZUMA (Center for Survey Research and Methodology, Mannheim, Germany). Weighting was done prior to the statistical analyses and was completed in two steps. First, the design was weighted using the number of phone connections and target persons per household. Second, adaptation weighting was done on the basis of the German Microcensus 2004 data according to the variables age, gender and education. Following the weighting of the data, the frequency distributions for age, gender and education corresponded to the Microcensus data for this age group in Baden-Wuerttemberg.

Procedures

Approval of the Ethics Committee of the Medical Faculty of Heidelberg was obtained and all participants consented to take part in the study. The developed questionnaire was pilot tested. Forty-seven subjects with an age, gender and education level distribution representative of that of the elderly in the study region were interviewed. The questionnaire was tested with respect to applicability, comprehensibility, completeness of the answer possibilities and length of the interview. No problems regarding applicability, comprehensibility and answer completeness were noted in this phase. Minor revisions to the questionnaire to reduce the interview length were undertaken.

The study participants were then surveyed from May to October 2006 by trained staff of a university phone lab (Chemnitz University of Technology), using the standardized questionnaire. All telephone interviews were computer-assisted (CATI). On average, each interview lasted 33 min. The questionnaire employed was integrated into 'The Survey System' software (Creative Research Systems Petaluma, CA 94952, USA), which enabled simultaneous data acquisition and storage and thus avoided any transfer errors.

The sampling of subjects followed the basic principle of the Gabler-Haeder method, which also includes phone numbers not listed in public telephone directories.[22] The target subjects were identified on the basis of a two-stage selection process. A telephone number was selected from the number pool using a random algorithm and the household contacted by phone. The 50- to 70-year-old target person with the most recent birthday was then interviewed for the study.

At the beginning of the interview, the subjects were informed about the purpose of the survey, the voluntary nature of participation and the anonymity of the processed data. Data were collected in an anonymous manner without disclosing the subject's identity, and transferred blinded to the authors. More detailed information on the study design and methods is given in Becker and colleagues.[23]

Measurements and study operationalization

The questionnaire used in this survey was based on validated, field-tested questionnaires from the German National Health Survey (BGS 1998),[24] the German National Telephone Health Survey (GSTel 2003)[25] and the Social Monitoring of Trends in Attitudes and Behaviours in Germany Survey (Allbus 2004).[26] Specific questions were created according to the research interests of the study. Field experts were consulted and asked to evaluate and optimise the questionnaire.

Four dimensions were identified: regular tobacco use, excessive alcohol consumption, an unhealthy diet and physical inactivity were included in the questionnaire, as the most important behavioural risk factors for morbidity and mortality.[27] These four factors have been considered in most other comparable studies.[7,14,16,18–21]

Regular tobacco use

Tobacco use was assessed with the following question: 'Do you currently smoke?' Possible answers were 'yes, daily', 'yes, occasionally' and 'no'. Subjects were also asked whether they had ever smoked. The number of cigarettes consumed per day was also investigated among current and former smokers. For the statistical analyses, a dummy variable was created for regular tobacco use, coded as 1 for daily consumption irrespective of tobacco type and 0 for otherwise.

Excessive alcohol consumption

All study participants were asked to state the amount of beer, wine and spirits they had consumed over the past weekend and on the last day of the week. For operationalization purposes, this survey used the German Nutrition Society upper limit for safe alcohol consumption of 10 g per day for women and 20 g per day for men.[28] These limits are also in accordance with the Ministry of Health recommendations.[29] Alcohol consumption exceeding the upper limit of 10 g and 20 g, respectively, was coded as 1 and 0 otherwise.

Unhealthy diet

Eating habits were documented on the basis of the consumption frequency of various foods. The categories were (cold) meats, dairy products, fish, fresh and frozen vegetables, fruit, deep-fried foods, whole grain products, white-flour products as well as confectionery and baked goods. During the interview subjects were asked if they had consumed foods from each category daily, several times a week, once a week, less often or never during the last 12 months.

An unhealthy diet was defined as those containing less than the daily recommended servings of either fresh/frozen fruit or vegetables. This definition is based on the German Nutrition Society recommendation of at least 250 g of fruit and vegetables per day as part of a healthy diet.[28]

Physical inactivity

Subjects were asked about the nature, duration, frequency and intensity of current, regular sporting activities. The 'Living an Active Life' study used an extensive and detailed questionnaire to investigate physical activity by type of sporting activity. Using Weitkunat's system,[21] individuals were defined as physically inactive if they did not perform at least one sporting activity on a regular basis for at least 1 h per week for a whole year at the time of the interview.

Socioeconomic status

Sociodemographic and socioeconomic variables were documented according to standardized methods.[30] Socioeconomic status (SES) was assessed using an additive index comprising income, education and occupational status according to Winkler's method.[31]

Details of the four indicators of health behaviour (regular tobacco use, excessive alcohol consumption, unhealthy diet and physical inactivity), demographic factors and SES are summarized in table 1.

Table 1 Reference values and coding for health behaviour and sociodemographic characteristics

	Coding
Health behaviour	
Regular tobacco use	1 = daily tobacco use (cigarettes, cigars, pipe and other smoking tobacco products)
Excessive alcohol consumption	Men: 1 = 20 g or more alcohol/day Women: 1 = 10 g or more alcohol/day
Unhealthy diet	1 = non-daily consumption of fresh or deep-frozen fruits or vegetables
Physical inactivity	1 = all year long continuous <1 h of exercise per week
Sociodemographic characteristics	
Socioeconomic status	0 = lower class 1 = middle and upper class
Gender	0 = male 1 = female
Age	0 = 50–59 years 1 = 60–70 years
Partnership status	0=currently living together with a partner (married or unmarried) 1 = currently without a partnership (widowed, separated, divorced or never-married, individuals living alone)

Statistical analyses

Cluster analysis was used to identify distinct health behaviour clusters. This multivariate analysis can be useful for finding homogeneous subgroups within heterogeneous samples.[32] The procedure employed was in accordance with the most recent developments in cluster analysis.[33] As the precise number of identifiable clusters was not known *a priori*, Ward agglomerative hierarchical clustering was used as it is particularly suitable for binary data.[34,35] First, the Ward method treated each individual observation as its own cluster. These clusters were gradually agglomerated to one large cluster on the basis of a proximity measure using a predefined fusion algorithm.[32] To enable identification of robust groups of observations, the fusion algorithm was stopped at the point where the individual clusters were as homogenous as possible within clusters and as heterogeneous as possible in relation to all the other clusters.[34,36] The established measures R^2, semi-partial R^2, pseudo F and pseudo t^2-statistics were used as the criteria for decisions regarding the total number of clusters. Finally, root mean square standard deviation (RMSSTD) was calculated as a measure of homogeneity.

A *post hoc* analysis looked at whether the groups identified by cluster analysis could be characterized on the basis of members' specific social attributes. A multinomial logistic regression with stepwise selection was used for this purpose.[37] Therefore, the odds of cluster membership were modelled for one social attribute while the other factors were held constant. The coefficients thus calculated can be interpreted as changes in the membership probability of the analysed cluster vs. the reference category (Cluster 1).

In accordance with standard statistical procedure, bivariate and multivariate analyses were only done on full datasets ($n = 1889$). All tests were two-tailed at a level of significance of $P \leq 0.05$. The analyses were conducted using the statistical program SAS 9.1®.

Results

Prevalence of unfavourable health behaviours in the population

One in seven of the over 50 year olds surveyed were currently daily smokers (15%). The percentage of smokers was typically higher in men than in women (18% vs. 12%). Current smokers consumed an average of 15 cigarettes daily (men: 16, women: 14 cigarettes per day). Two out of three respondents (67%) said they occasionally drank alcohol (men: 74%, women: 61%). Twenty-nine percent of subjects reported that they did not eat fresh/frozen fruit or vegetables daily (men: 31%, women: 27%). A total of 50% of respondents said they were physically inactive (men: 53%, women: 46%).

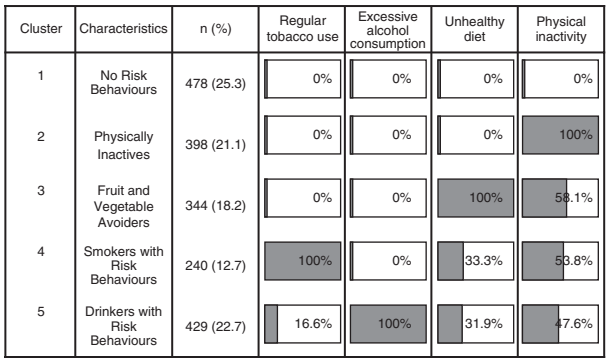

Figure 1 Distribution of health behaviour indicators in the clusters

Health behaviour clusters

Cluster analysis identified five distinct groups with typical patterns of health behaviour. All measures derived from the analysis showed a five-cluster pattern in the form of a definite change in the graphic fusion curve. Results of the RMSSTD also supported this five-cluster pattern.

A total of 1889 subjects provided information on all health behaviour variables and could therefore be included in the cluster analysis. Of these, 25.3% were categorized as having 'No Risk Behaviours', 21.1% as 'Physically Inactives', 18.2% as 'Fruit and Vegetable Avoiders', 12.7% as 'Smokers with Risk Behaviours' and 22.7% as 'Drinkers with Risk Behaviours'. The clusters are described by the respective predominant behaviour in more detail as follows (figure 1).

Cluster 1: 'No Risk Behaviours'

One quarter of all the individuals included in the sample reported consuming no alcohol or tobacco, having a healthy diet and engaging regularly in sports. No one in this cluster had unfavourable health behaviour as measured by the above described indicators. This cluster contained the largest number of study participants.

Cluster 2: 'Physically Inactives'

The second cluster identified was comprised of individuals whose only health-harming behaviour was insufficient physical activity but who were without other risk behaviours. None of the individuals in this group regularly used tobacco or excessively consumed alcohol and all of them had a healthy diet. The 'Physically Inactives' were the third largest cluster in numerical terms.

Cluster 3: 'Fruit and Vegetable Avoiders'

All individuals in this group reported an unhealthy diet, the key indicator of this cluster. Moreover, six out of 10 individuals in this cluster did not get enough exercise.

Table 2 Results of the multinomial logistic regression for sociodemographic characteristics

	Cluster 2: 'Physically Inactives'	Cluster 3: 'Fruit and Vegetable Avoiders'	Cluster 4: 'Smokers with Risk Behaviours'	Cluster 5: 'Drinkers with Risk Behaviours'
Socioeconomic status				
Middle and upper class vs. lower class	ns	ns	0.35 (0.17–0.72)	ns
Gender				
Female vs. male	ns	0.61 (0.42–0.88)	0.37 (0.25–0.55)	0.60 (0.43–0.84)
Age				
60–70 years vs. 50–59 years	ns	ns	0.37 (0.25–0.53)	ns
Partnership status				
Currently without a partner vs. with partner	1.60 (1.11–2.31)	1.98 (1.34–2.91)	2.62 (1.70–3.96)	ns

Odds ratio (95% confidence interval)
Reference category: Cluster 1 ('No Risk Behaviours')
Model fit: Likelihood-ratio test: $\chi^2 = 103.98$; df=20; Pseudo $R^2 = 0.0223$
ns, not significant

Cluster 4: 'Smokers with Risk Behaviours'

The smallest cluster was comprised entirely of smokers. Smoking was combined with inadequate physical activity and/or an unhealthy diet in 75% of the cases. Along with the 'Drinkers with Risk Behaviours' described below, members of this cluster displayed the highest average number of combined unhealthy behaviours.

Cluster 5: 'Drinkers with Risk Behaviours'

The second largest cluster was centred on excessive alcohol use. In 66% of all cases, high alcohol consumption occurred in conjunction with at least one other unhealthy behaviour. Hence, this cluster also included a large number of combined unhealthy behaviours. This group also included all individuals with unfavourable values in all four health behaviour clusters. Each of the above clusters represents the sequential accumulation of health-harming behaviours.

Sociodemographic characteristics of the clusters

To facilitate specific target group prevention measures, sociodemographic indicators of the clusters were investigated and set in relation to Cluster 1 ('No Risk Behaviours'). The odds ratios for cluster membership according to different social factors are shown in table 2. The composition of the 'health-aware' Cluster 1 was as follows: 60.4% female, 46.3% aged 50–59 and 71% in the middle or upper SES categories.

In contrast with the reference group, the 'Physically Inactives' (Cluster 2) were significantly more likely to live alone (OR = 1.60). The 'Fruit and Vegetable Avoiders' (Cluster 3) were also significantly more likely to be without a partner (OR = 1.98). In particular, this group was comprised of divorced or widowed men.

The 'Smokers with Risk Behaviours' (Cluster 4) were less likely to be female (OR = 0.37), to be within the oldest age group (OR = 0.37) and to have middle or high SES (OR = 0.35). This cluster largely included younger men with low SES, indicating an interaction effect between gender and SES (OR = 3.41; 95% CI: 1.13–10.28). This group also was more likely to live alone (OR = 2.62).

As with the previous two clusters (Clusters 3 and 4), the 'Drinkers with Risk Behaviours' were significantly less likely to be female (OR = 0.60) and also showed an interaction effect between gender and SES (OR = 2.95; 95% CI: 1.36–6.39).

The health behaviour clusters identified were plotted on a two-dimensional matrix by gender and SES. Figure 2 shows the size and position of the clusters, which was based on the proportion of men and individuals in the upper social class. Generally, the higher the percentage of men in a cluster, the more unfavourable the health behaviour. The interaction effect between gender and SES is clearly demonstrated for Cluster 4 ('Smokers with Risk Behaviours').

Discussion

Principal findings

Among the over 50 year olds, we identified five clusters of typical health behaviour. One group represented the ideal type of health behaviour in terms of risk and prevention and two groups were associated with regular use of tobacco and excessive consumption of alcohol, commonly in conjunction with several other risk behaviours. The other clusters were a mixture of both favourable and unfavourable health behaviour patterns.

Relation to other studies

International studies on this topic cover much wider age ranges (from 18 and 20 to 59 and 84 years).[8,12–14,16–20] These studies also demonstrate that there is a high correlation between behavioural risk factors. In some studies, 20–30% of the participants engaged in least three out of the four, often even all four, risk behaviours.[15,18,19] This phenomenon is usually more pronounced in men.

Although later studies contrasted the same individual health behaviours, clusters were not generated (except in the case of Abel).[7,8,17,19] Using US and German cross-sectional data, he identified three clusters of health behaviour including smoking, alcohol consumption, physical activity and diet. However, the analysis was less differentiated and comprised an earlier sample with a large age range between 18 and 84 years.

The few studies which actually identified multidimensional clusters are based on much smaller samples or non-representative cohorts.[6,20,21] These studies were conducted in the 1990s in German-speaking countries and were based on data from the cities Munich (Germany), Bern (Switzerland) and the Frankfurt region (Germany). Interestingly, the health behaviour clusters identified in those studies show striking similarities to those identified in this study. All three publications describe one fairly large cluster associated with no health-harming behaviours. The cluster with physical inactivity as the sole health hazard observed in this study

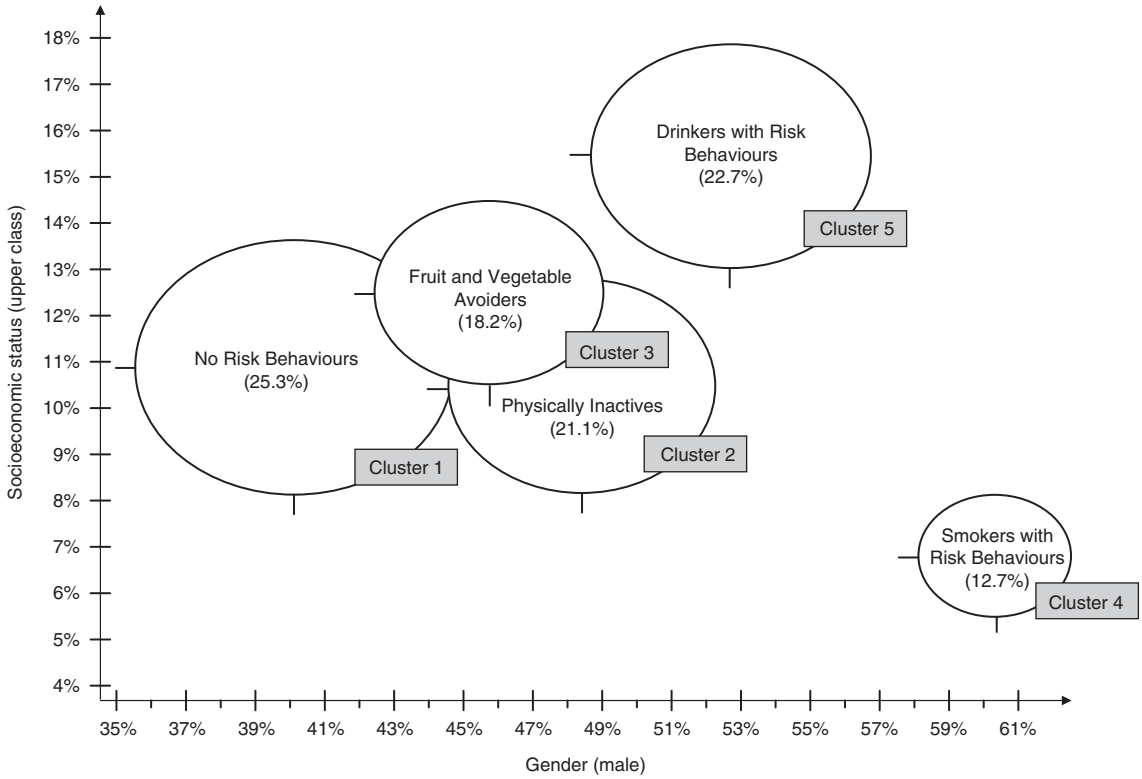

Figure 2 Cluster description on the basis of gender and socioeconomic status

('Physically Inactives') and the unhealthy diet cluster ('Fruit and Vegetable Avoiders') were also identified in all three of the abovementioned studies. The two health behaviours smoking and drinking were broken down into two to three clusters in previous studies.

Methodological considerations

The design of the 'Living an Active Life' study is cross-sectional. Therefore, we can only provide a snapshot of the current health behaviour of the participants. However, we hypothesize that our data reflects typical behaviour patterns. At least as far as German-speaking areas are concerned, health behaviour clusters seem to remain stable over time. In this study, we particularly emphasized the regularity of physical activities, asked about current and former smoking habits, and interviewed participants about their eating habits over the last 12 months.

One main limitation arising out of the nature of the study is that the data generated reflect information elicited from the subjects themselves. As such, social desirability may be an issue.[38] While the proportion of subjects who engage in sporting activities on a regular basis is obviously higher in this study as compared to others,[39,40] this percentage is in accordance with the results of the GSTel 2003 study.[25] The variability in different studies indicates that not only social desirability but also the type of question, the context and the range of possible answers might also play a role.[23]

Because the subjects included in the study all resided in Baden-Wuerttemberg it is impossible to decide whether the data are representative for Germany. Additionally, the hierarchical cluster method does not allow for the inclusion of weighting factors. This also affects the generalizability of statements about the absolute size of clusters, although the different health behavioural clusters themselves have been validated.

With respect to analyses of the descriptive aspects of the study, the application of weighting factors was possible. In this case the prevalence for the four health behaviour indicators correspond well with national values from official public health sources.[41]

This study's strength lies in the novel and innovative application of a method used in other scientific fields. Thus, it is the first study which has identified a typology of health behaviour patterns in the over 50-year-old population in the German-speaking area by cluster analysis.

Implications and future research

Health-harming behaviours evidently occur in very typical combinations, which can be taken into account when developing effective and efficient prevention and intervention measures. For instance, there are two high-risk groups among older adults: individuals who drink too much alcohol on a regular basis (who tend to be men with higher SES) and those who are regular smokers (who tend to be younger seniors with lower SES). Both risk groups exhibit specific sociodemographic attributes (male, living alone, social class affiliation) and behavioural patterns (physical inactivity and an unhealthy diet). Another target group is the 'Physically Inactives', who tend to have a lower SES and live alone. This cluster can be expected to be responsive to an entirely different target group approach.

There is evidence that combinations of these four important and most prevalent risk factors are more detrimental to people's health than would be expected from the addition of the individual effects alone.[18] This multiplicative rather than additive effect suggests a need for a multimodal approach such as is the established gold standard in the diagnosis and management of many diseases.[42] For example, in addiction therapy, current discussions have centred on treating alcohol and tobacco dependency at the same time and including exercise programmes as part of therapy programmes. To the

extent that future research work identifies similarly stable and complex health behaviour types for other countries and age groups, this would open up an additional opportunity to implement larger scale prevention strategies.

In the face of the demographical change, there will have to be more specific programmes and interventions for younger seniors in the future. The basis of this recommendation is the typical cumulative risk patterns observed among the elderly. Individual and societal lifestyle changes can also be worthwhile among the elderly.[43]

Acknowledgements

'Living an Active Life' is a representative survey of middle-aged and older adults, centred on individual sports and health biographies, health behaviour and lifestyles. Finally, the authors acknowledge the support of the Landesstiftung Baden-Wuerttemberg (September 2005 to August 2007) for this study. The authors also extend their appreciation to Shelby Yamamoto for providing language support.

Conflicts of interest: None declared.

Key points

- Many prevention and intervention measures are still targeting isolated behaviours like tobacco smoking or physical inactivity.
- This study used cluster analysis for the identification of homogenous health behaviour patterns among target groups for the purpose of identifying preventive measures and interventions in the over 50-year-old population.
- The following clusters were identified: 'No Risk Behaviours', 'Physically Inactives', 'Fruit and Vegetable Avoiders', 'Smokers with Risk Behaviours' and 'Drinkers with Risk Behaviours'.
- Whereas the first cluster is the lowest in terms of risk and the ideal with respect to prevention, the last two clusters, regular tobacco use and excessive consumption of alcohol, are associated with additional risk factors.

References

1 World Health Organisation. *Diet, physical activity and health. Executive Board 109th Session, provisional agenda item 3.10, EB109/14*. Geneva: World Health Organisation, 2001.

2 World Health Organisation. *The World Health Report 1995. Bridging the gaps*. Geneva: World Health Organisation, 1995.

3 Steptoe A, Wardle J. What the experts think: a European survey of expert opinion about the influence of lifestyle on health. *Eur J Epidemiol* 1994;10:195–203.

4 Wynder EL. Listen to nature. The challenge of lifestyle medicine. *Soz Präventivmed* 1991;36:137–46.

5 Kickbusch I. Life-styles and health. *Soc Sci Med* 1986;22:117–24.

6 Spegel H, Weitkunat R, Crispin A, et al. Muster des Gesundheitsverhaltens und ihre Determinanten - Ergebnisse des Bern-Münchner Lebensstilpanels (Patterns and determinants of health behaviour - Results of the Bern-Munich Lifestyle Panel). *Soz Präventivmed* 1999;44:160–70.

7 Abel T. Konzept und Messung gesundheitsrelevanter Lebensstile (Concept and measuring of health lifestyles). *Prävention* 1992;15:123–8.

8 Abel T. Measuring health lifestyles in a comparative analysis: theoretical issues and empirical findings. *Soc Sci Med* 1991;32:899–908.

9 Bailey KD. Sociological classification and cluster analysis. *Qual Quant* 1983;17:251–68.

10 Anderberg MR. *Cluster analysis for applications*. New York: Academic Press, 1973.

11 Klastorin TD. Assessing cluster analysis results. *J Mark Res* 1983;20:92–8.

12 Bracco MM, Colugnati FA, Pratt M, Taddei JA. Multivariate hierarchical model for physical inactivity among public school children. *J Pediatr* 2006;82:302–7.

13 Faeh D, Viswanathan B, Chiolero A, et al. Clustering of smoking, alcohol drinking and cannabis use in adolescents in a rapidly developing country. *BMC Public Health* 2006;6:169.

14 Chiolero A, Wietlisbach V, Ruffieux C, et al. Clustering of risk behaviors with cigarette consumption: a population-based survey. *Prev Med* 2006;42:348–53.

15 Coups EJ, Gaba A, Orleans CT. Physician screening for multiple behavioral health risk factors. *Am J Prev Med* 2004;27(2 Suppl):34–41.

16 Ma J, Betts NM, Hampl JS. Clustering of lifestyle behaviors: the relationship between cigarette smoking, alcohol consumption, and dietary intake. *Am J Health Promot* 2000;15:107–17.

17 Padrao P, Lunet N, Santos AC, Barros H. Smoking, alcohol, and dietary choices: evidence from the Portuguese National Health Survey. *BMC Public Health* 2007;7:138.

18 Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev Med* 2007;44:124–8.

19 Schuit AJ, van Loon AJ, Tijhuis M, Ocke M. Clustering of lifestyle risk factors in a general adult population. *Prev Med* 2002;35:219–24.

20 Opper E. *Sport – Ein Instrument zur Gesundheitsförderung für alle? Eine empirische Untersuchung zum Zusammenhang von sportlicher Aktivität, sozialer Lage und Gesundheit* (Exercise - a health promoting instrument for all? An empirical study of the association of sporting activity, social position and health). Aachen: Meyer & Meyer, 1998.

21 Weitkunat R. *Computergestützte Telefoninterviews als Instrument der sozial- und verhaltensepidemiologischen Gesundheitsforschung* (Computer assisted telephone interviews as instrument of the health research in social and behavioural epidemiology). Berlin: Logos, 1998.

22 Gabler S, Häder S. Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland (Advisements towards a sample design for telephone interviews in Germany). *ZUMA-Nachr* 1997;41:7–18.

23 Becker S, Huy C, Brinkhoff KP, et al. ''Ein aktives Leben leben'' - Sport, Bewegung und Gesundheit im mittleren und höheren Erwachsenenalter. Konzeption, Datenerhebung und erste Ergebnisse eines repräsentativen Basis-Survey für die 50- bis 70-jährige baden-württembergische Wohnbevölkerung ("Living an active life" - sports, exercise and health in middle-aged and older adults. An empirical database on physical activity, health behaviour and lifestyle in the 50- to 70-year-old residential population of Baden-Wuerttemberg). *Gesundheitswesen* 2007;69:401–7.

24 Bellach BM, Knopf H, Thefeld W. Der Bundes-Gesundheitssurvey 1997/98 (The German Health Survey 1997/98). *Gesundheitswesen* 1998;60 (Suppl 2):S59–68.

25 Kohler M, Rieck A, Borch S, Ziese T. *Erster telefonischer Gesundheitssurvey des Robert Koch-Institus. Beiträge zur Gesundheitsberichterstattung des Bundes* (First National Telephone Health Survey by Robert Koch-Institute. Articles for the health reporting system of the Federal Republic of Germany). Berlin: Robert Koch-Institute, 2005.

26 German Social Science Infrastructure Services. *Allbus 2004. Fragebogen ZA-Nr. 3752* (Allbus 2004. Questionnaire ZA-No. 3752). Cologne: Gesis, 2004.

27 Lopez AD, Mathers CD, Ezzati M, et al. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* 2006;367:1747–57.

28 German Nutrition Society, Austrian Nutrition Society, Swiss Nutrition Society, Swiss Society of Nutrition Research, Swiss Nutrition Association. *Referenzwerte für die Nährstoffzufuhr* (Reference values for nutrient intake). Frankfurt/Main: Umschau/Braus, 2000.

29 Burger M, Brönstrup A, Pietrzik K. Alkoholkonsum und Krankheiten. *Schriftenreihe des Bundesministeriums für Gesundheit* (Alcohol consumption

and diseases. Series of the German Federal Ministry of Health). Baden-Baden: Nomos, 2000.

30 Working Committee of the German Institutes for Marketing and Social Research, Working Group of the Socioscientific Institutes, German Federal Statistical Office. *Demographische Standards* (Demographical standards). Wiesbaden: German Federal Statistical Office, 2004.

31 Stolzenberg H. Bundes-Gesundheitssurvey 1998. Dokumentation des Datensatzes (German National Health Survey 1998. Documentation of the dataset). *Public Use File BGS98*. Berlin: Robert Koch-Institute, 2000.

32 Backhaus K, Erichson B, Plinke W, Weiber R. *Multivariate Analysemethoden (Multivariate analysis methods)*. Berlin: Springer, 2006.

33 Milligan GW, Sokal L. A two-stage clustering algorithm with robustness recovery characteristics. *Educ Psychol Meas* 1980;40:755–9.

34 Finch H. Comparison of distance measures in cluster analysis with dichotomous data. *J Data Sci* 2005;3:85–100.

35 Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963;58:236–44.

36 Dillon WR, Goldstein M. *Multivariate analysis: methods and applications*. New York: Wiley, 1984.

37 Long JS. *Regression models for categorical and limited dependent variables*. Thousand Oaks: Sage, 2000.

38 Latenschlager GJ, Flaherty VL. Computer administration of questions: more desirable or more social desirability. *J Appl Psychol* 1990;75:310–4.

39 Caspersen CJ, Pereira MA, Curran KM. Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Med Sci Sports Exerc* 2000;32:1601–9.

40 Mensink GB. Körperliche Aktivität (Physical activity). *Gesundheitswesen* 1999;61:S126–31.

41 Ellert U, Wirz J, Ziese T. *Telefonischer Gesundheitssurvey des Robert Koch-Instituts, 2. Welle* (National Telephone Health Survey by Robert Koch-Institute, 2nd wave). Berlin: Robert Koch-Institute, 2006.

42 Albert SM. *Public health and aging: an introduction to maximizing function and well-being*. New York: Springer, 2004.

43 World Health Organisation. Active ageing. *A policy framework*. Geneva: World Health Organisation, 2002.

Received 9 July 2008, accepted 22 December 2008