Dietary patterns and lifestyle

Their determinants and associations with chronic disease burden

PhD Thesis, Utrecht University, the Netherlands-with a summary in Dutch

ISBN:	978-90-393-6475-8
Author:	Heidi Fransen
Coverdesign:	Hilde Wolters
Lay-out:	Heidi Fransen & Hilde Wolters
Print:	Ipskamp Printing

The research reported in this thesis was financially supported by the Netherlands Organisation for Health Research and Development (ZonMw, grant no 40-00812-98-10040). Printing of this thesis was financially supported by the Julius Center for Health Sciences and Primary Care.

Dietary patterns and lifestyle

Their determinants and associations with chronic disease burden

Voedingspatronen en leefstijl

Determinanten en associaties met ziektelast

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 16 februari 2016 des middags te 2.30 uur

door

Heidi Petra Fransen
geboren op 30 maart 1978
te Venray

Promotor: prof. dr. P.H.M. Peeters

Copromotoren: dr. A.M. May
dr. ir. J.W.J. Beulens

Contents

Chapter 1

General introduction

GENERAL INTRODUCTION

A long life in good health. That is what most people hope for. In recent decades, life expectancy in the Netherlands has increased between 1970 and 2014 from 76.5 to 83.3 years for women and from 70.8 to 79.9 years for men (1). Cancer and cardiovascular disease are examples of non-communicable or chronic diseases. Although worldwide the number of premature deaths due to non-communicable diseases decreases, the disease burden of these diseases increased, from 43% in 1990 to 54% in 2010 (2). This increase is caused by the increased number of years people live with disabilities (3). In Western Europe approximately 60% of the total disease burden is due to chronic diseases (2). Unhealthy lifestyle factors are important risk factors for chronic diseases. Diet, smoking behavior, weight and physical activity are modifiable lifestyle factors: changing to the healthier variant of these lifestyle factors may result in a longer life in good health (4).

Traditionally, research on diet and chronic disease risk focused on the association between one specific nutritional determinant and one disease outcome of interest. However, in real life many nutritional components are consumed together, resulting in a dietary pattern. Dietary patterns in their turn relate to risk of not one but several chronic diseases. In this thesis the association between several dietary patterns and the total burden of chronic diseases is studied. Quality-adjusted life years are used as a measure of overall disease burden. The ultimate goal of this project is to identify the healthiest dietary pattern in the Netherlands. In addition, it also aims to provide insight in lifestyle factors that are associated with this pattern in the Netherlands. Furthermore, determinants of other unhealthy lifestyle characteristics are studied.

DIETARY PATTERNS

Dietary patterns can be defined in two ways: a priori (based on current knowledge) and a posteriori (data driven). In this thesis both methods are used. We selected five dietary patterns that were most relevant for the Dutch population: three a priori defined patterns: the Mediterranean Diet, the Healthy Diet according to the WHO, and the Dutch Healthy Diet based on guidelines defined by the Dutch Health Council. Two a posteriori defined patterns (a 'Western' and a 'prudent' pattern) are based on data of a large Dutch cohort, the EPIC-NL cohort (5). Throughout this thesis data of the EPIC-NL cohort is used, which consists of the 2 Dutch cohorts that are included in the European Prospective Investigation into Cancer and Nutrition (EPIC) study (6, 7) (text box 1).

In this thesis, adherence to a Mediterranean diet was assessed by the modified Mediterranean Diet Score (mMDS), developed by Trichopoulou et al. (8). This score includes nine dietary components: vegetables; legumes; fruit, nuts and seeds; cereals; fish; the ratio of unsaturated to saturated fatty acids; meat; dairy products; and alcohol. Median intake levels of the Dutch EPIC-NL study population were used as a cut-off value to score intake of individuals. Intake equal to or above the median intake was assigned 1 point, except for meat and dairy products. For meat and dairy, intake equal to or below the median was assigned a value of 1. Alcohol consumption (10-50 g alcohol per day for men or 5-25 g per day for women) was assigned a value of 1. So, never alcohol consumers, former drinkers and excessive drinkers were assigned zero points. The mMDS score ranges from 0 (minimal adherence) to 9 (maximal adherence). It should be noted that EPIC-NL participants with a high adherence to this dietary pattern (high mMDS) on average do not compare to people living in the South of Europe. Southern Europeans have on average much higher absolute intake levels of above products. Therefore throughout this thesis we will refer to this diet as a Mediterranean-style diet.

The Healthy Diet Indicator (HDI) is based on WHO recommendations for the prevention of chronic diseases (9) and includes six nutrients (saturated fatty acids, polyunsaturated fatty acids, cholesterol, protein, dietary fiber, and free sugars) and one food group (fruits and vegetables combined). To compute the indicator, individual intake of each component is compared with the WHO recommended intake.
The Dutch Healthy Diet index (DHD-index) is based on the Dutch dietary guidelines for a healthy diet, defined by the Dutch health council in 2006 (10). Eight dietary components are included in this continuous score: vegetables, fruit, fiber, fish, saturated fatty acids, trans fatty acids, sodium and alcohol.

Principal Component analysis identified two main a posteriori patterns in the EPIC-NL cohort and, based on identified components, they were labeled 'Western' and 'prudent' (5). The 'Western' dietary pattern represents a high intake of French fries, fast food, low-fiber products and soft drinks, while the 'prudent' pattern represents a high intake of fish, vegetables and high-fiber products.

LIFESTYLE FACTORS

In addition to unhealthy diet, other unhealthy lifestyle factors such as smoking, being overweight and physical inactivity, are related to chronic disease occurrence. The association of modifiable lifestyle factors with either mortality or morbidity has been

studied frequently (11-14), also in the EPIC-NL population (15, 16). In this thesis we study associations with overall disease burden.

QUALITY-ADJUSTED LIFE YEARS

To express overall disease burden in a population, summary health measures can be computed, such as quality-adjusted life years (QALY) or disability-adjusted life years (DALY). Both combine information of morbidity and mortality from several diseases into one outcome measure. They provide a comprehensive measure of the relation between risk factors and overall disease burden.

A QALY combines information on life expectancy and quality of life. Years lived with a disease are weighed with a 'utility weight' that represents the reduction in quality of life attributable to having a specific chronic disease. One QALY thus equals one year in optimal health, while zero QALY equals death. QALYs were originally developed for use in economic evaluations of health interventions to measure health gain when comparing interventions (17). In this thesis, we compute QALYs and DALYs for participants of the EPIC-NL cohort and relate them to dietary patterns and other lifestyle patterns. In previous studies, we related lifestyle and dietary patterns to DALYs (18, 19).

DETERMINANTS OF AN UNHEALTHY DIET AND LIFESTYLE

Socio-demographic and lifestyle factors have been related to an (un)healthy diet before. A low educational level, a low income or a low occupational position have been associated with an unhealthy diet (20), while a higher educational level or occupational position have been associated with a healthier diet (21). Physical inactivity, smoking and a young age were also related to an unhealthy diet (22, 23). Furthermore, unhealthy behaviors tend to cluster in persons with a low socio-economic status (24, 25). It is not clear if unhealthy behaviors co-exist to the same extent in high educated people. This information can be useful in the development of targeted health promotion strategies.

OUTLINE OF THIS THESIS

In this thesis we study five dietary patterns: three patterns that are based on current knowledge (a priori patterns) and two data driven patterns (a posteriori). Several subjective decisions are used to determine the number of dietary patterns to retain in a posteriori dietary pattern analysis. As different dietary pattern solutions can vary in food group composition, this may affect reported associations with disease outcome. In the first part of this thesis, ***methodology***, we therefore examined reliability of different pattern solutions, 2 to 6 patterns, and applied quantitative criteria to determine the number of patterns to extract (**chapter 2**).

In the second part we investigate the associations of ***lifestyle and diet*** with disease burden (**chapter 3-5**). Modifiable lifestyle factors such as diet, smoking, being overweight and physical inactivity, are related to chronic disease occurrence. Their association with mortality or morbidity has been studied before, but in this thesis we relate them to overall disease burden. In **chapter 3** four modifiable lifestyle factors, smoking behavior, weight, physical activity level and diet are related to disease burden, using quality-adjusted life years (QALYs). In addition, we combine the four lifestyle factors into a healthy lifestyle score and investigate the association between this healthy lifestyle score and QALYs. In **chapter 4** we relate five dietary patterns to QALYs to investigate which dietary pattern is associated with the lowest burden of chronic diseases in the Netherlands. It is still heavily debated if alcohol consumption should be part of a healthy dietary pattern and included in public health guidelines. Moderate alcohol consumption has been associated with a reduced risk of cardiovascular disease, but an increased risk of some cancers. Therefore, in this thesis, we investigated the association between alcohol consumption and overall disease burden, using DALYs (**chapter 5**).

In the final part ***determinants of an unhealthy diet and lifestyle*** are reported. In **chapter 6** we investigate whether associations between unhealthy lifestyle factors and unhealthy diets differ between people with a low and a high educational level. Finally, in **chapter 7** we investigate if a relatively short period of transient undernutrition early in life is associated with an unhealthy lifestyle later in life. For this study, we used data of women who were exposed to the Dutch famine during childhood or adolescence. The results of this thesis are discussed in **chapter 8** and the main findings are presented in the **summary**.

Text box 1: the EPIC-NL study

EPIC-MORGEN & EPIC-Prospect

- 22,654 men and women
- aged 20-59 years
- from Doetinchem, Amsterdam, and Maastricht
- random sample of the Dutch population

- 17,357 women
- aged 49-70 years
- from Utrecht and vicinity
- participated in the nationwide breast cancer screening program

EPIC-NL study

- Dutch contribution to European Prospective Investigation into Cancer and Nutrition (EPIC)
- combination of EPIC-MORGEN and EPIC- Prospect

- 40,011 men and women
- recruited in 1993-1997
- aged 20-70 years at recruitment

Data collection at recruitment:
- physical examination
- lifestyle questionnaire, a.o. education, previous illnesses, smoking, alcohol consumption, and physical activity level

- food frequency questionnaire asking about habitual frequency and portion size of consumption of foods the year preceding enrolment: 178 food items

- prospectively followed for occurrence of incident diseases and death by linkage to registries, until December 2007

Reference: Beulens et al., 2010
Cohort Profile: The EPIC-NL Study.
Int J Epidemiol 2010 Oct;39(5):1170-8

REFERENCES

1. Statistics Netherlands. Levensverwachting; geslacht en leeftijd, vanaf 1950 (per jaar). Accessed September 2015. Available at: http://statline.cbs.nl.

2. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012; 380 (9859): 2197-223.

3. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369(10):954-64.

4. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859): 2224-60.

5. Stricker MD, Onland-Moret NC, Boer JM, van der Schouw YT, Verschuren WM, May AM, et al. Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke. Nutr Metab Cardiovasc Dis. 2013 Mar;23(3):250-6.

6. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8.

7. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14.

8. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocke MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ. 2005;330(7498):991.

9. Huijbregts P, Feskens E, Rasanen L, Fidanza F, Nissinen A, Menotti A, et al. Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and The Netherlands: longitudinal cohort study. BMJ. 1997;315(7099):13-7.

10. van Lee L, Geelen A, van HE, de Vries JH, Van't Veer P, Feskens EJ. The Dutch Healthy Diet index (DHD-index): an instrument to measure adherence to the Dutch Guidelines for a Healthy Diet. Nutr J. 2012;11:49.

11. van den Brandt PA. The impact of a Mediterranean diet and healthy lifestyle on premature mortality in men and women. Am J Clin Nutr. 2011;94(3):913-20.

12. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med. 2012;55(3):163-70.

13. Kvaavik E, Batty GD, Ursin G, Huxley R, Gale CR. Influence of individual and combined health behaviors on total and cause-specific mortality in men and women: the United Kingdom health and lifestyle survey. Arch Intern Med. 2010;170(8):711-8.

14. Knoops KT, de Groot LC, Kromhout D, Perrin AE, Moreiras-Varela O, Menotti A, et al. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. JAMA. 2004;292(12):1433-9.

15. Struijk EA, May AM, Beulens JW, Fransen HP, de Wit GA, Boer JM, et al. Adherence to the Dutch Guidelines for a Healthy Diet and cancer risk in the European Prospective Investigation into Cancer and Nutrition-Netherlands (EPIC-NL) cohort. Public Health Nutr. 2014;17(11):2546-53.

16. Struijk EA, May AM, Wezenbeek NL, Fransen HP, Soedamah-Muthu SS, Geelen A, et al. Adherence to dietary guidelines and cardiovascular disease risk in the EPIC-NL cohort. Int J Cardiol. 2014;176(2):354-9.

17. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96:5-21.

18. Struijk EA, Beulens JW, May AM, Fransen HP, Boer JM, de Wit GA, et al. Dietary patterns in relation to disease burden expressed in Disability-Adjusted Life Years. Am J Clin Nutr. 2014;100(4):1158-65.

19. May AM, Struijk EA, Fransen HP, Onland-Moret NC, de Wit GA, Boer JM, et al. The impact of a healthy lifestyle on Disability-Adjusted Life Years: a prospective cohort study. BMC Med. 2015;13:39.

20. Kant AK. Dietary patterns and health outcomes. J Am Diet Assoc. 2004; 104(4):615-35.

21. Boylan S, Lallukka T, Lahelma E, Pikhart H, Malyutina S, Pajak A, et al. Socio-economic circumstances and food habits in Eastern, Central and Western European populations. Public Health Nutr. 2011; 14(4):678-87.

22. Hu EA, Toledo E, Diez-Espino J, Estruch R, Corella D, Salas-Salvado J, et al. Lifestyles and risk factors associated with adherence to the Mediterranean diet: a baseline assessment of the PREDIMED trial. PLoS One. 2013;8(4):e60166.

23. Patino-Alonso MC, Recio-Rodriguez JI, Belio JF, Colominas-Garrido R, Lema-Bartolome J, Arranz AG, et al. Factors associated with adherence to the Mediterranean diet in the adult population. J Acad Nutr Diet. 2014;114(4):583-9.

24. Mullie P, Clarys P, Hulens M, Vansant G. Dietary patterns and socioeconomic position. Eur J Clin Nutr. 2010; 64(3):231-8.

25. Pampel FC, Krueger PM, Denney JT. Socioeconomic Disparities in Health Behaviors. Annu Rev Sociol. 2010;36:349-70.

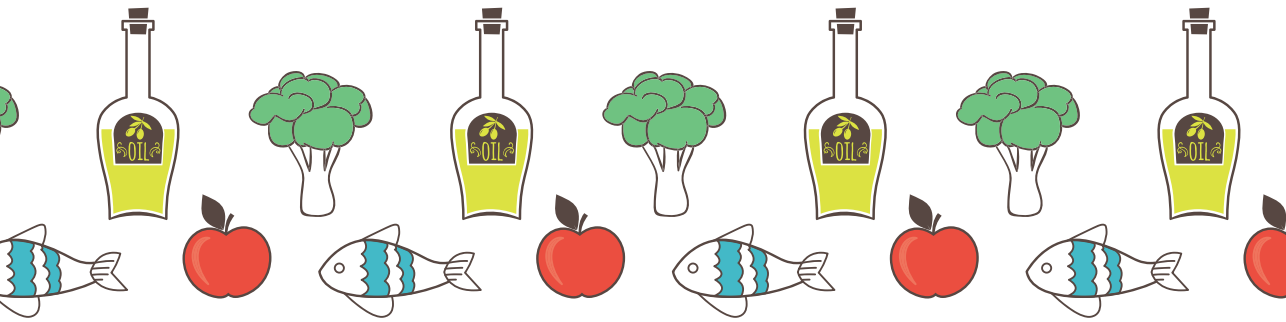
Methodology

Chapter 2

A posteriori dietary patterns: how many patterns to retain?

Based on: Heidi P Fransen, Anne M May, Martin D Stricker, Jolanda M Boer, Christian Hennig, Yves Rosseel, Marga C Ocké, Petra HM Peeters, Joline WJ Beulens.
A posteriori dietary patterns: how many patterns to retain?

J Nutr. 2014 Aug;144(8):1274-82.
doi: 10.3945/jn.113.188680



ABSTRACT

Principal component analysis (PCA) and cluster analysis are used frequently to derive dietary patterns. Decisions on how many patterns to extract are primarily based on subjective criteria, whereas different solutions vary in their food-group composition and perhaps association with disease outcome. Literature on reliability of dietary patterns is scarce, and previous studies validated only 1 preselected solution. Therefore, we assessed reliability of different pattern solutions ranging from 2 to 6 patterns, derived from the aforementioned methods. A validated food frequency questionnaire was administered at baseline (1993-1997) to 39,678 participants in the European Prospective Investigation into Cancer and Nutrition-The Netherlands (EPIC-NL) cohort. Food items were grouped into 31 food groups for dietary pattern analysis. The cohort was randomly divided into 2 halves, and dietary pattern solutions derived in 1 sample through PCA were replicated through confirmatory factor analysis in sample 2. For cluster analysis, cluster stability and split-half reproducibility were assessed for various solutions. With PCA, we found the 3-component solution to be best replicated, although all solutions contained ≥1 poorly confirmed component. No quantitative criterion was in agreement with the results. Associations with disease outcome (coronary heart disease) differed between the component solutions. For all cluster solutions, stability was excellent and deviations between samples was negligible, indicating good reproducibility. All quantitative criteria identified the 2-cluster solution as optimal. Associations with disease outcome were comparable for different cluster solutions. In conclusion, reliability of obtained dietary patterns differed considerably for different solutions using PCA, whereas cluster analysis derived generally stable, reproducible clusters across different solutions. Quantitative criteria for determining the number of patterns to retain were valuable for cluster analysis but not for PCA. Associations with disease risk were influenced by the number of patterns that are retained, especially when using PCA. Therefore, studies on associations between dietary patterns and disease risk should report reasons to choose the number of retained patterns.

INTRODUCTION

A wealth of literature investigating the relation of dietary patterns with all kinds of clinical outcomes has been published since this field emerged. Dietary patterns can be derived either a priori or a posteriori (1). While a priori techniques use scoring systems to assess the degree to which a participant complies with a predefined dietary pattern, a posteriori methods derive patterns empirically based on the observed dietary intake using principal component analysis (PCA), exploratory factor analysis (EFA), or cluster analysis. PCA derives patterns ("components") by grouping highly correlated food groups using data reduction (2). PCA is most commonly used for dietary pattern analysis (1), although EFA may theoretically be more applicable because it takes common variance in observed variables into account whereas PCA only considers total variance. When defining patterns, subjective decisions are introduced at various points, such as decisions for cutoffs for food-group loadings or type of rotation (1, 3). Furthermore, the Kaiser criterion, scree plot, and interpretability, i.e., criteria that are either subjective and/or tend to over-extract (4), are mostly used to determine the number of patterns to retain (1).

In contrast to PCA, cluster analysis is not based on the correlation between food groups but clusters individuals. It identifies groups of participants, i.e., "clusters", who share the same eating habits based on differences in food intakes (5). There are numerous clustering algorithms, but in dietary pattern analysis, Ward's cluster analysis and K-means cluster analysis (KCA) are most commonly used (1). To determine the number of patterns to retain in KCA, researchers usually request different cluster solutions and select the best interpretable solution (1). Hence, the selection of the number of patterns to retain is based on subjective criteria for both methods. No quantitative criteria are being used currently. Meanwhile, dietary patterns based on different solutions usually vary in their food-group composition, which may affect associations with disease outcome. Furthermore, reliability may differ considerably between different solutions. Previous studies examined reliability or reproducibility of mostly 1 preselected solution derived from PCA, EFA or cluster analysis (6-10). Only 1 study replicated several pattern solutions to determine the optimal number of clusters to extract (11).

Using the European Prospective Investigation into Cancer and Nutrition-The Netherlands (EPIC-NL) cohort, we examined and compared reliability of different solutions ranging from 2 to 6 patterns derived from PCA and cluster analysis. We applied quantitative criteria new to this field to determine the number of patterns to extract and explored their agreement with the reliability results. Furthermore, we investigated differences in

the ability of different pattern solutions to predict disease outcome by investigating associations with coronary heart disease (CHD) incidence during follow-up.

PARTICIPANTS AND METHODS

Study population

EPIC-NL consists of the Prospect-EPIC (17,357 women aged 50-69 y) and the Monitoring Project on Risk Factors for Chronic Diseases (MORGEN)-EPIC (22,654 men and women aged 20-64 y) cohorts. All participants were recruited between 1993 and 1997 through random population sampling in Amsterdam, Maastricht and Doetinchem (MORGEN-EPIC) or after participation in the breast cancer screening program (Prospect-EPIC). Written informed consent was obtained before study inclusion (12). At baseline, a FFQ and a general questionnaire were filled out. The study complies with the Declaration of Helsinki and was approved by the Institutional Review Board of the University Medical Center Utrecht for Prospect-EPIC and the Medical Ethics Committee of the Netherlands Organization for Applied Scientific Research for MORGEN-EPIC (12).

We excluded 333 participants because of missing dietary information or extreme energy intake (<600 or >5000 kcal/d). The remaining 39,678 were randomly half-split into a derivation and a replication sample. For the association with CHD incidence, 4101 participants were additionally excluded because of the presence of myocardial infarction (n=724) or stroke (n=448) at baseline or loss to follow-up (n=2613), resulting in 35,910 participants to study this association.

Assessment of dietary intake

To assess dietary intake, a validated FFQ containing questions on the habitual frequency of consumption of 79 food items during the year preceding enrollment was used. Additional information was sought on consumption frequency for different sub-items, preparation methods, and additions, which ultimately allowed the estimation of consumption of 178 food items (13, 14). For each of the foods, intake was calculated in grams per day by multiplying the consumption frequency with the portion size, which was either estimated using standard weights or provided by the participants (color photographs). Total energy (kilojoules per day) intake was then estimated using an extended version of the 1996 computerized Dutch food composition table (15). The FFQ was validated for food groups (and nutrients) with 12 monthly 24-h recalls and biomarkers in 24-h urine and serum samples. The median relative validity (Pearson's

correlations) was 0.61 for men and 0.53 for women (13, 14). To identify dietary patterns, we grouped the food items of the FFQ into 31 food groups (Supplemental Table 1), as described previously (16).

Ascertainment of CHD

Participants were followed for the occurrence of diseases through linkage with the hospital discharge diagnoses database from the Dutch National Medical Registry, using a validated probabilistic procedure (17). For participants who died during follow-up, information on the cause of death was obtained through linkage with the Cause of Death Registry from Statistics Netherlands. Both fatal and nonfatal CHD events were included (International Classification of Diseases 9th edition: 410-414, 427.5, 798.1, 798.2, 798.9; International Classification of Diseases 10th edition: I20-I25, I46, R96) (12). When multiple events of CHD occurred, the first diagnosis was taken as endpoint. Follow-up ended on the day of diagnosis, day of death, or at the end of the study (December 31, 2007).

Statistical methods

Treatment of food group variables

For each of the 31 food groups, intakes were represented as the percentage energy they contributed to total energy intake (18). For PCA and EFA, these food-group variables (in energy percentage) were directly used. For KCA, food-group variables were top coded, i.e., values that were higher/lower than the mean $\pm$ 6 times the SD were changed to the value of the mean $\pm$ 6 times the SD, because KCA is sensitive to outliers (1). Furthermore, no additional standardization was applied because comparisons of using standardized or unstandardized variables suggested that it gives undue influence to minor food groups (1).

PCA

Component solutions from PCA were confirmed using confirmatory factor analysis (CFA). For CFA, a hypothesis is needed on the food patterns to include in the analysis (2). Here, the components that were derived from PCA were used. With CFA, the retained patterns are tested to investigate which solution best fits the underlying observed data. A scheme of analysis is presented in Figure 1. PCA with varimax rotation was used to extract solutions ranging from 2 to 6 factors in the derivation sample (principal function in R-package "psych") (19). As shown in Figure 1 (arrow a), each of the obtained solutions was then separately confirmed in the replication sample using CFA (cfa

function; lavaan) (20, 21). Additionally, the procedure was repeated using EFA as detailed in Supplemental Methods 1. Food groups with factor loadings in PCA (FL_{PCA}) of $\geq|0.25|$ were included in the models for CFA. Food groups with factor loadings in CFA (FL_{CFA}) of $\geq|0.20|$ were retained. To summarize the results, the ratio of food groups not confirmed (n food groups with $FL_{CFA} \leq |0.20|$) to the total number of included food groups (n food groups with $FL_{PCA} \geq |0.25|$) and the deviation in factor loadings between PCA and CFA, defined as

$$Deviation(PCA-CFA) = \frac{1}{k} \cdot \sum_{i=1}^{k} 100 \cdot |(Fl_{PCA} - Fl_{CFA})/Fl_{PCA}|$$

(where k= n food groups), were computed (Table 1). For example, a value of 20% means that FL_{CFA} of a particular food group differed on average 20% from FL_{PCA}.

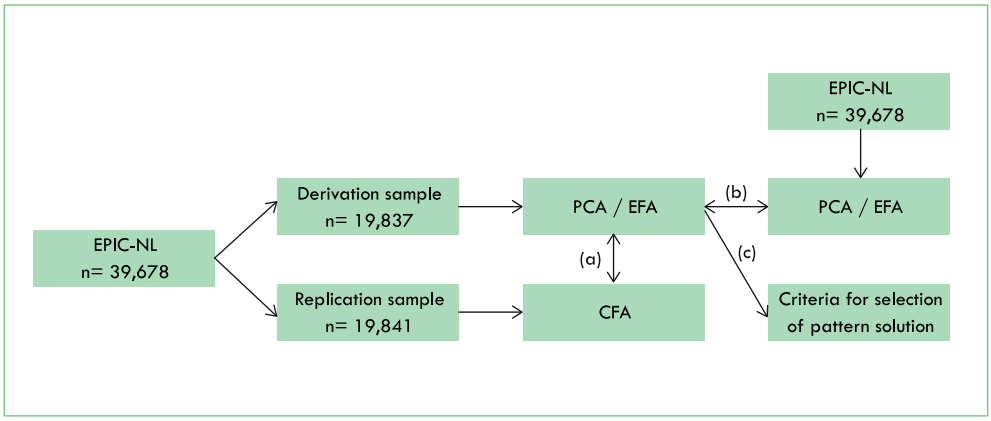


Figure 1 PCA and EFA: scheme of statistical analyses in the EPIC-NL cohort. Arrow (a) shows a comparison between PCA/EFA (derivation sample) and CFA (replication sample). Arrow (b) is a comparison between PCA/EFA (derivation sample) and PCA/EFA (whole study population). Arrow (c) shows how the Kaiser criterion, scree test optimal coordinate, and visual inspection of the scree plot were applied to identify the number of factors/components to extract.

Arrow b in Figure 1 shows how internal reproducibility was checked by performing PCA in the replication sample and the entire study population. The Kaiser criterion, the scree test optimal coordinate, and a visual inspection of the scree plot (the point at which the slope of the graph changes) were applied to identify the number of patterns to retain (Fig. 1, arrow c). Details are provided in Supplemental Methods 2. Besides reliability of the component solutions, the interpretability of the dietary patterns is also an important tool. We also checked whether the solutions we find are well interpretable.

Cluster analysis

For cluster analysis, a similar procedure was followed as for PCA (Fig. 2). Five cluster solutions ranging from 2 to 6 clusters were extracted in both subsamples and the entire study population to examine reproducibility (Fig. 2, arrows a and b). In Figure 2 (c), cluster stability was measured for all solutions extracted in the derivation sample, using Jaccard similarities. Finally, to quantitatively identify the number of patterns to retain, internal cluster validity indices were calculated, i.e., the Calinski-Harabasz and Davies-Bouldin indices (Fig. 2, arrow d). Details of these indices are provided in Supplemental Methods 3.

Using KCA, 5 different cluster solutions ranging from 2 to 6 clusters were extracted from both samples (the function *kmeans* in the R-package stats) (5, 22). Within each cluster solution, each food group was assigned to the cluster where it contributed the highest mean intake. To measure internal reproducibility, deviations between the results obtained in the derivation and replication samples were computed, defined as

$$Deviation(CA) = \frac{1}{k} \cdot \sum_{i=1}^{k} 100 \cdot | (x_{0i} - x_{1i}) / x_{0i} |$$

where k= n food groups, x0*i* = the mean intake of food group *i* in the derivation sample, and x1*i* = the mean intake of food group *i* in the replication sample. For example, a value of 2% means that, on average, the mean intake of a particular food group in a particular cluster varies 2% between the derivation and the replication samples. The clusters in the derivation and replication samples were matched manually. Data analysis was performed with R version 2.15.2 (22).

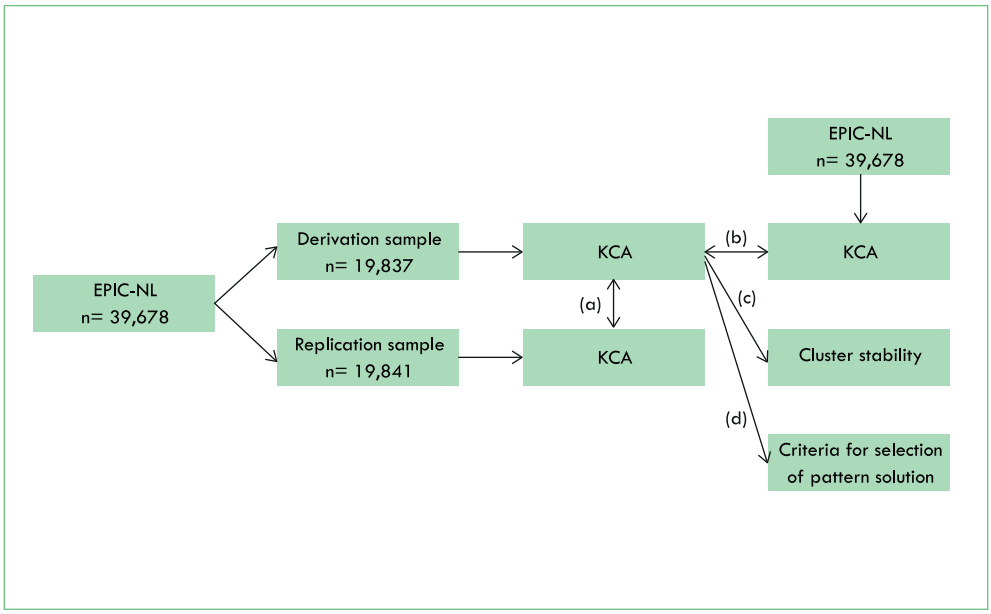


Figure 2 Cluster analysis: scheme of statistical analyses in the EPIC-NL cohort.
Arrow (a) shows internal reproducibility, i.e., comparison between the derivation and replication samples. Arrow (b) is a comparison between the derivation sample and the whole study population. Arrow (c) shows cluster stability (Jaccard similiarities). Arrow (d) shows internal cluster validity indices to evaluate the optimal cluster solution. EPIC-NL, European Prospective Investigation Into Cancer and Nutrition-The Netherlands; KCA, K-means cluster analysis.

Associations between dietary patterns and CHD

Cox proportional hazard analysis was used to ascertain associations between dietary patterns and CHD risk. The analysis was performed for all component and cluster solutions. Quartiles of the components scores (PCA) or cluster membership were used as the independent variable and incident CHD as the dependent variable. HRs and 95% CIs are presented. Results are adjusted for the following: *1)* age; *2)* gender; *3)* physical activity (inactive, moderately inactive, moderately active, or active); *4)* smoking status and intensity (never; former, defined as quit smoking >20 y ago, quit 10-20 y ago, quit ≤10 y ago; or current, defined as <15 cigarettes/d, 15-25 cigarettes/d, >25 cigarettes/d, or pipe or cigar smoker); *5)* education (low, average, or high); and *6)* energy intake. For these analyses, IBM SPSS Statistics for Windows version 20 (IBM) was used. *P* values <0.05 were considered statistically significant.

Table 1 Confirmation success of different component solutions obtained by principal component analysis in the European Prospective Investigation into Cancer and Nutrition-The Netherlands cohort (n=39,678).

Pattern solution and pattern	Food groups not confirmed/total number of food groups	Deviation	Mean deviation
2			27.8
A	2/14	18.7	
B	7/14	37.0	
3			24.5
A	1/13	17.8	
B	5/11	36.0	
C	1/11	19.7	
4			33.0
A	4/10	29.8	
B	0/10	23.3	
C	2/11	23.3	
D	9/11	55.5	
5			28.5
A	0/5	14.7	
B	7/9	48.7	
C	2/12	27.5	
D	1/10	27.9	
E	2/9	23.6	
6			26.0
A	0/5	10.8	
B	2/11	29.5	
C	0/9	20.7	
D	4/6	31.5	
E	1/6	35.1	
F	0/8	28.0	

RESULTS

Demographic and lifestyle characteristics of the 2 half-split samples were almost identical (Supplemental Table 2).

PCA

Detailed food-group composition of all solutions is shown in Table 2. Overall, different component solutions contained patterns with different food-group compositions. In the 2-component solution, a more Western dietary pattern containing high loadings on French fries, fast food, and soft drinks and a more prudent pattern with fish, vegetables and high-fiber products were found. In the 3-component solution, the prudent pattern was subdivided into 2 patterns. When component 4 was extracted, the more Western dietary pattern was also subdivided into 2 patterns (Table 2). Finally, when replicated in the whole study population or in the other split half, PCA led to comparable food-group loadings for all patterns as for the split-half sample (data not shown).

Confirmation success of all component solutions is presented in Table 1. The 2 measures for the reliability of the patterns were well in line with each other, i.e., high ratios of food groups not confirmed to the total number of food groups were mostly accompanied by high deviations in food-group loadings between PCA and CFA.

Confirmation success differed considerably between patterns within the same solution. For example, component 2A was better confirmed than component 2B, i.e., the ratio of food groups not confirmed to the total number of food groups (2 of 14 vs. 7 of 14) and the deviation in food-group loadings between PCA and CFA were much lower (18.7% vs. 37.0%). Overall, the 3-component solution was better confirmed than the other component solutions (mean deviation of 24.5%), although all solutions contained ≥1 poorly confirmed pattern with a deviation >30% (2B, 3B, 4D, 5B, 6D, and 6E). Results of quantitative criteria to determine the number of patterns to retain are presented in Figure 3. After visual inspection of the scree plot, a qualitative criterion, one might select 3 patterns, because the slope of the graph changes at 3 solutions. The scree test optimal coordinates suggested extracting 8 patterns. The Kaiser criterion recommended selecting even more patterns (11 components). As for PCA, reliability of dietary patterns differed considerably between different EFA solutions, and the scree plot was the only criterion indicating the same solution as the reliability measures: the 3-factor solution. Detailed results of the EFA are presented in Supplemental Methods 1 and Supplemental Tables 3 and 4.

Table 2 Food-group composition of component solutions 2-6 obtained by principal component analysis in the European Prospective Investigation into Cancer and Nutrition-The Netherlands cohort, n=39,678[1]

Pattern solution and pattern	Positively scoring food groups	Negatively scoring food groups
2		
A	French fries, fast food[2], soft drinks with sugar, low-fiber bread, low-fiber cereals, other alcoholic drinks[3]	Fruit, low-fat dairy products, high-fiber bread, cakes/cookies, high-fat dairy products[3], boiled vegetables/legumes, cheese, raw vegetables
B	Shellfish, high-fat fish, raw vegetables, low-fat fish, wine, low-fiber cereals[3], high-fiber cereals[3], fruit[3], fruit juices[3], chicken[3]	Potatoes, fat/butter[3], sugar/sweets, low-fiber bread[3]
3		
A	French fries, fast food[2], soft drinks with sugar, low-fiber bread, low-fiber cereals, other alcoholic drinks[3]	Low-fat dairy products, fruit, high-fiber bread, cakes/cookies, high-fat dairy products, boiled vegetables/legumes, cheese
B	Shellfish, high-fat fish, low-fat fish, raw vegetables, wine, chicken[3], boiled vegetables/legumes[3], eggs[3], fruit[3]	Sugar/sweets, high-fat dairy products[3]
C	High-fiber cereals, nuts[4], raw vegetables, low-fiber cereals, fast food[2]	Potatoes, red meat, processed meat, fat/butter, low-fiber bread, boiled vegetables/legumes[3]
4		
A	Shellfish, low-fat fish, high-fat fish, raw vegetables, chicken[3], boiled vegetables/ legumes[3], fruit[3], wine, eggs[3]	Sugar/sweets
B	French fries, fast food[2], other alcoholic drinks, processed meat, low-fiber cereals, soft-drinks with sugar	Low-fat dairy products, high-fat dairy products, fruit, cakes/cookies
C	High-fiber cereals, nuts[4], low-fiber cereals, raw vegetables, fast food[2]	Potatoes, red meat, fat/butter, processed meat, boiled vegetables/legumes[3], low-fiber bread[3]
D	Low-fiber bread, soft drinks with sugar[3], French fries[3], sugar/ sweets[3], fast food[2,3], other non-alcoholic drinks[3]	High-fiber bread, cheese[3], wine[3], fruit[3], raw vegetables[3]
5		
A	Shellfish, high-fat fish, low-fat fish, raw vegetables, wine	
B	High-fiber bread, raw vegetables[3], fruit[3], wine[3], cheese[3], boiled vegetables /legumes[3]	Low-fiber bread, sugar/sweets[3], fat/butter[3]

Table 2 continued

Pattern solution and pattern	Positively scoring food groups	Negatively scoring food groups
5		
C	Low-fat dairy products, high-fat dairy products, fruit, high-fiber cereals, cakes/cookies, fruit juices[3], boiled vegetables/ legumes	Other alcoholic drinks, processed meat, French fries, fast food[2], wine[3]
D	Soft drinks with sugar, other non-alcoholic drinks, French fries, low-fiber cereals, fast food[2], chicken[3]	Cheese, cakes/cookies, fat/butter, high-fiber bread
E	Red meat, potatoes, boiled vegetables/legumes, processed meat, chicken[3]	Nuts[4], high-fiber cereals, fast food[2], sugar/sweets[3]
6		
A	Shellfish, high-fat fish, low-fat fish, raw vegetables, wine	
B	Low-fat dairy product, high-fat dairy products, fruit, cakes/cookies, high-fiber cereals, fruit juices[3], boiled vegetables/legumes	Other alcoholic drinks, processed meat, French fries, fast food[2], wine[3]
C	Potatoes, red meat, processed meat, fat/butter, boiled vegetables/legumes	Nuts[4], high-fiber cereals, low-fiber cereals, fast food[2]
D	High-fiber bread, cheese[3], raw vegetables[3], fruit[3]	Low-fiber bread, sugar/sweets[3]
E	Soft drinks with sugar, other non-alcoholic drinks, French fries, fast food[2]	Cheese[3], cakes/cookies
F	Oils and margarine, chicken, raw vegetables, boiled vegetables/legumes, low-fiber cereals, red meat	Fat/butter, sugar/sweets

[1] Food groups are sorted in descending order according to component loadings.
[2] Contains spring rolls, Russian salad, pizza and Dutch fried meat snacks.
[3] Not confirmed food groups, i.e., food groups with food-group loadings < |0.20| in confirmatory factor analysis.
[4] Contains nuts, seeds, soy products, and peanut butter.

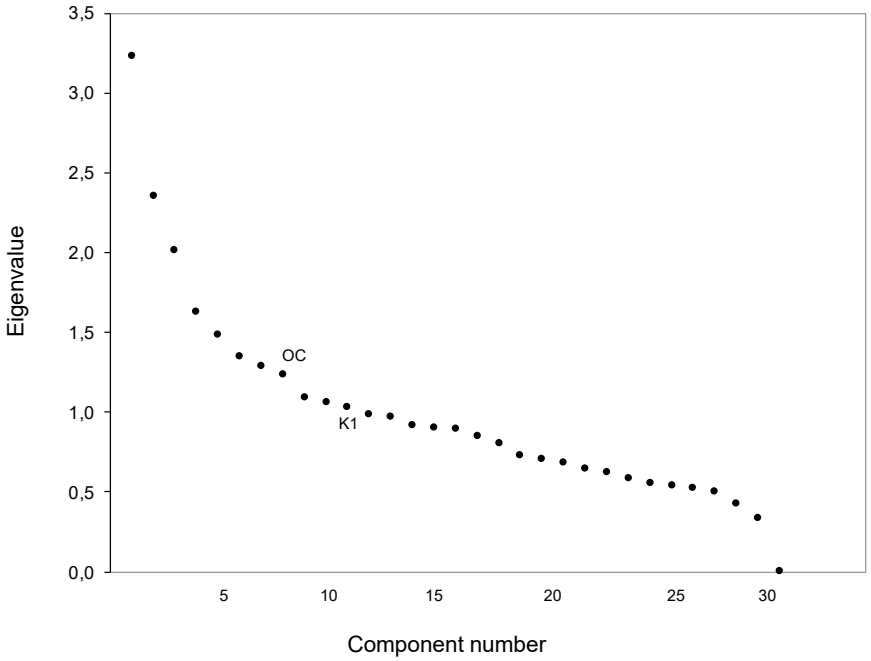


Figure 3 Scree plot and results from scree test optimal coordinates (OC) and Kaiser criterion (K1). Observed eigenvalues after PCA and corresponding eigenvalues obtained from uncorrelated normal variables are plotted against the component number. PCA, principal component analysis.

Associations with disease outcome (CHD) differed between the various component solutions (Table 3). For the 2-component solution, the prudent pattern (2B) tended to be associated with a lower risk of CHD (HR for extreme quartiles: 0.90; 95% CI: 0.78, 1.04), wheras the Western pattern was not associated with CHD (HR: 0.93; 95% CI: 0.79, 1.10). The Western pattern from the 3-component solution was not associated with CHD (HR: 0.95; 95% CI: 0.80, 1.13), whereas only 1 of the 2 prudent patterns (3C) remained associated with a lower risk of CHD (HR: 0.80; 95% CI: 0.69, 0.93). In the 4-component solution, 1 of the Western patterns (4D) became associated with an increased CHD risk (HR: 1.25; 95% CI: 1.09, 1.44), whereas the prudent pattern remained associated with a lower CHD risk (HR: 0.79; 95% CI: 0.68, 0.92). For the 5-component solution, none of the 3 prudent patterns were associated with a lower CHD risk, but both Western patterns were associated with an increased risk. In the 6-component solution, no significantly lower risk on CHD was found in any of the patterns, whereas 2 Western patterns remained associated with an increased CHD risk (6C and 6E).

Table 3 Incident CHD for dietary patterns extracted through principal component analysis in the European Prospective Investigation Into Cancer and Nutrition-The Netherlands (n=35,910)[1]

	Quartile			
Dietary pattern	1	2	3	4
2A	Reference	0.88 (0.77, 1.00)	0.99 (0.86, 1.14)	0.93 (0.79, 1.10)
2B	Reference	0.96 (0.85, 1.09)	0.93 (0.82, 1.06)	0.90 (0.78, 1.04)
3A	Reference	0.89 (0.78, 1.01)	1.01 (0.87, 1.16)	0.95 (0.80, 1.13)
3B	Reference	1.10 (0.97, 1.26)	1.01 (0.88, 1.15)	1.02 (0.89, 1.18)
3C	Reference	0.96 (0.85, 1.08)	0.87 (0.77, 0.99)	0.80 (0.69, 0.93)
4A	Reference	1.03 (0.90, 1.17)	0.97 (0.84, 1.11)	1.13 (0.99, 1.30)
4B	Reference	1.00 (0.87, 1.15)	1.04 (0.90, 1.20)	1.05 (0.91, 1.22)
4C	Reference	0.90 (0.80, 1.01)	0.85 (0.75, 0.97)	0.79 (0.68, 0.92)
4D	Reference	1.16 (1.02, 1.32)	1.23 (1.08, 1.40)	1.25 (1.09, 1.44)
5A	Reference	0.95 (0.83, 1.08)	0.98 (0.86, 1.11)	0.94 (0.82, 1.08)
5B	Reference	1.06 (0.93, 1.20)	0.93 (0.82, 1.06)	0.90 (0.78, 1.03)
5C	Reference	1.01 (0.88, 1.15)	0.99 (0.86, 1.14)	1.07 (0.92, 1.24)
5D	Reference	1.19 (1.06, 1.35)	1.09 (0.96, 1.25)	1.32 (1.15, 1.53)
5E	Reference	1.08 (0.93, 1.27)	1.17 (1.01, 1.36)	1.38 (1.19, 1.60)
6A	Reference	0.89 (0.78, 1.02)	0.96 (0.85, 1.10)	0.92 (0.81, 1.05)
6B	Reference	1.04 (0.91, 1.19)	0.99 (0.86, 1.14)	1.05 (0.91, 1.22)
6C	Reference	1.07 (0.92, 1.25)	1.09 (0.93, 1.27)	1.35 (1.16, 1.57)
6D	Reference	0.95 (0.84, 1.08)	0.87 (0.76, 0.99)	0.94 (0.83, 1.07)
6E	Reference	1.25 (1.10, 1.42)	1.28 (1.12, 1.46)	1.43 (1.24, 1.65)
6F	Reference	1.04 (0.92, 1.18)	1.04 (0.91, 1.18)	1.04 (0.91, 1.19)

[1] Values are HRs (95% CIs). Adjusted for age, gender, physical activity, smoking status and intensity, education, and energy intake. CHD, coronary heart disease.

Cluster analysis

Food-group compositions and results for cluster stability and reproducibility are presented in Table 4. All clusters of all solutions obtained Jaccard similarities >0.85, indicating that they were highly stable. However, the 2-cluster solution had the best overall stability with Jaccard similarities of 1.00. All clusters were well reproduced, i.e., deviations between the split-half samples were small, although values steadily increased from the 2-cluster (1.1%) to the 6-cluster (4.3%) solution. These findings are well in line with results from the internal cluster validity indices, which consistently indicated the 2-cluster solution as optimal (Table 5). Cluster analysis performed in the whole study population led to almost identical clusters for all solutions as for the 2 split-half samples (data not shown). Regarding the interpretability of the cluster solutions we found, the first 5 solutions all clearly had 1 more prudent dietary pattern that included fish, high-fiber products, vegetables and fruit (patterns 2A, 3C, 4B, 5B, and 6E). The more Western dietary pattern obtained for the 2-cluster solution (2B) was subdivided into different clusters when more solutions were retained.

The more prudent pattern obtained from the 2-cluster solution was associated with a lower risk of CHD compared with the more Western pattern (HR: 0.91; 95% CI: 0.82, 1.00) (Table 6). For the other solutions, the pattern containing low-fiber bread, potatoes, fat, and butter was used as the reference. The lower risk of the more prudent patterns in the 3-, 4-, 5-, and 6-cluster solutions was comparable with the result obtained from the 2-cluster solution (HR values between 0.90 and 0.92). Surprisingly, patterns containing, among others, red meat and alcoholic drinks were also associated with a lower risk (HR values between 0.76 and 0.85).

Table 4 Food-group composition, cluster stability, and internal reproducibility of all cluster solutions derived in the derivation sample of the European Prospective Investigation Into Cancer and Nutrition-The Netherlands cohort (n=39,678).

Pattern solution and Pattern	Cluster solution and patterns	Stability[1]	Deviation[2]	Mean deviation[3]
2A	Fish/shellfish, fruit juice, wine, high-fiber products, vegetables, fruits, nuts[4], cakes/cookies, low-fat dairy products, cheese	1.00	1.18	1.1
2B	Red/processed meat, alcoholic drinks other than wine, sugar products, low-fiber products, potatoes, French fries, fast food[5], fat/butter	1.00	0.97	
3A	High protein (chicken, meat, fish, eggs), alcoholic drinks (wine and other), low-fiber cereals, French fries, fast food[5],	0.87	2.69	1.8
3B	Low-fiber bread, potatoes, fat/butter, sugar/sweets	0.97	1.27	
3C	Low-fat fish, high-fiber products, vegetables, fruits, nuts[4], fruit juices, cakes /cookies, dairy products	0.98	1.36	
4A	Low-fiber bread, potatoes, fat/butter	0.95	2.68	2.5
4B	Low-fat fish, high-fiber products, vegetables, fruits, cakes/cookies, low-fat dairy products, cheese	0.98	1.47	
4C	Chicken, fruit juices, soft drinks with sugar, cereals, nuts[4], French fries, fast food[5], high-fat dairy products, sugar/ sweets	0.91	3.44	
4D	High protein (meat, fish, eggs), alcohol	0.96	2.40	
5A	Wine, high-fiber bread, cheese	0.90	2.91	3.4
5B	Fish, chicken, fruit juices, high-fiber cereals, vegetables, fruit, cakes/cookies, dairy products	0.87	4.37	
5C	Low-fiber bread, potatoes, fat/butter	0.96	2.59	
5D	Red/processed meat, eggs, alcoholic drinks other than wine	0.93	2.91	
5E	Chicken, sugar products, low-fiber products, nuts[4], French fries, fast food[5]	0.90	3.99	
6A	Shellfish, wine, high-fiber bread, nuts[4], oils/margarine, cheese	0.89	5.64	4.3
6B	Processed meat, low-fiber bread, potatoes, fat/butter	0.93	2.17	
6C	Red meat, high-fat fish, shellfish, eggs, other alcoholic drinks	0.93	3.63	
6D	No food group contributed the highest mean intake to this cluster	0.88	2.56	
6E	Low-fat fish, shellfish, fruit juices, high-fiber products, boiled vegetables/legumes, raw vegetables, fruit, cakes/cookies, low/high-fat dairy products	0.89	8.11	
6F	Chicken, soft drinks with sugar, other alcoholic drinks, low-fiber cereals, French fries, fast food[5], sugar/sweets	0.88	3.47	

[1] Jaccard similarities: >0.85 indicates highly stable,; 0.6-0.75 indicates patterns, but it is unclear which points belong to which cluster; <0.5 indicates dissolved cluster; [2] Mean deviation of food groups between 2 subsamples; [3] Mean deviation per cluster solution ;[4] Contains nuts, seeds, soy products, and peanut butter; [5] Contains spring rolls, Russian salad, pizza, and Dutch fried meat snacks.

Table 5 Internal cluster validity indices for all cluster solutions derived in the derivation sample of the European Prospective Investigation Into Cancer and Nutrition-The Netherlands cohort (n=39,678).

	Index	
Cluster solution	Calinski-Harabasz[1]	Davies-Bouldin[2]
2	5286	1.94
3	3354	2.09
4	2667	2.50
5	2272	2.71
6	2006	2.78

[1] Cluster solution with the maximum value is considered optimal;
[2] Cluster solution with the minimum value is considered optimal.

Table 6 CHD for dietary patterns extracted through K-means cluster analysis in the European Prospective Investigation Into Cancer and Nutrition-The Netherlands cohort (n=35,910)[1]

	Pattern					
Pattern solution	A	B	C	D	E	F
2	0.91 (0.82, 1.00)	Reference				
3	0.85 (0.76, 0.96)	Reference	0.92 (0.83, 1.02)			
4	Reference	0.92 (0.82, 1.03)	0.98 (0.87, 1.12)	0.79 (0.67, 0.93)		
5	0.90 (0.79, 1.02)	0.90 (0.79, 1.03)	Reference	0.76 (0.64, 0.91)	0.95 (0.82, 1.10)	
6	0.91 (0.79, 1.05)	Reference	0.78 (0.65, 0.94)	0.92 (0.79, 1.07)	0.90 (0.77, 1.05)	0.94 (0.79, 1.13)

[1] Values are HRs (95% CIs). Adjusted for age, gender, physical activity, smoking status and intensity, education, and energy intake. CHD, coronary heart disease.

DISCUSSION

The number of patterns to extract in dietary pattern analysis is often based on subjective criteria. In this study, we applied quantitative criteria, such as the scree test optimal coordinates and internal cluster validity indices, to determine the number of patterns to extract and explored their agreement with reproducibility over the split samples. Reproducibility differed significantly between various pattern solutions for PCA, whereas cluster analysis extracted stable and reproducible clusters across

different solutions. The applied quantitative criteria were found to be valuable for cluster analysis but not for PCA. Furthermore, we showed that the number of retained patterns influences conclusions regarding the associations with disease outcome, especially when using PCA.

PCA

In our study, the reliability of the various pattern solutions differed considerably. The 3-component solution was more reproducible over split samples than the other component solutions, although all solutions contained ≥1 poorly confirmed pattern. A reason for this lack of reproducibility could be that variation in food intakes of individuals is not completely captured in dietary patterns. The explained variance of 6 components was 39% in our study, indicating that other determinants also play an important role in variance in dietary intake. Imperfect measurement by FFQs may also play a role. PCA analyzes the total variance assuming that the food groups are perfectly reproduced by the extracted components. As a result of measurement error in dietary intake (23) and the low-to-moderate correlations between food groups, this probably is not the case in dietary pattern analysis. Our results indicate that there is added value to first use split-half techniques and CFA as a tool to find the optimal number of components. PCA could subsequently be rerun over the whole population to obtain scores for the selected solution only. Interestingly, one may have opted for the 3-component solution after visual inspection of the scree plot as well, whereas both quantitative criteria, i.e., the Kaiser criterion and the scree test optimal coordinates, recommended extracting considerably more patterns. PCA aims to reduce the number of food groups into a small number of dietary patterns. Therefore, it is not reasonable to extract, for example, 11 patterns out of 31 food groups as suggested by the Kaiser criterion, whereas many of these patterns contain only a few high-scoring food groups. For these reasons, our results suggest that these quantitative criteria seem not to be useful in dietary pattern analysis. It also indicates that it remains important to consider interpretability of the obtained patterns when applying quantitative criteria to decide on the number of patterns to extract, because patterns with only a few high-scoring food groups may not represent meaningful dietary patterns. Moreover, because different pattern solutions contain patterns with different food-group compositions, the choice of the final solution affects the obtained associations with disease outcome. When more components were extracted, we still identified prudent and Western patterns, but the associations of the obtained prudent patterns with CHD incidence disappeared.

Cluster analysis

Because cluster analysis tends to give results even for fairly homogenous datasets, cluster validation is important (24). Based on cluster stability and internal reproducibility, KCA derived highly stable and reproducible clusters for all solutions. Internal cluster validity indices consistently indicated the 2-cluster solution as optimal, which is in line with results from analyses on stability and reproducibility. This solution clearly distinguished a prudent pattern from a more Western dietary pattern and was therefore well interpretable. This more distinctly prudent pattern was also found in the 3-, 4-, 5- and 6-cluster solutions, whereas the Western pattern was subdivided into different clusters with more solutions. The different number of retained patterns had less effect on associations with disease risk in cluster analysis than in PCA. The more distinctly prudent dietary patterns tended to suggest an inverse association with CHD risk in all solutions, but none reached statistical significance.

Comparison with literature

A direct comparison of our results with the literature is hampered by different study designs and research questions. Other studies replicated only 1 preselected solution. For PCA, Lau et al. (9) found a 2-component solution to be reproducible. However, they merely fitted the data because PCA and CFA were performed in the same sample (2). We used a derivation and validation set, and, consequently, differences in component/factor loadings between PCA and CFA were higher in our study, leading to nonconfirmed patterns. Newby et al. (6) also assessed reproducibility by performing PCA in random half-split samples. They found drastic differences in component loadings between the split halves and the whole population for some food groups, whereas we observed comparable loadings. These differences might well be explained by the larger population in our study (39,678 vs. 459).

For cluster analysis, only 2 studies examined internal reproducibility or stability of patterns, using either split-half samples (6) or discriminant analysis (8). Quatromoni et al. (8) extracted a 5-cluster solution through Ward's clustering and used discriminant analysis to measure stability and classification ability. In total, 80% of all participants were correctly classified, indicating good stability. Newby et al. (6) assessed reproducibility by performing KCA in random half-splits and found good reproducibility. Lo Siou et al. (11) investigated reproducibility for 3 different cluster methods (Ward's, flexible β, and KCA), and KCA was found to have the highest reproducibility. By examining the ratios of between-cluster and within-cluster variances, comparable with the Calinski-Harabasz-index, they chose the number of clusters to

retain and found the 4-cluster solution to be optimal in men. In women, no optimal solution could be found using this method, because of the little variation in intake.
Our results are in line with these studies showing highly stable and reproducible clusters obtained by KCA, but we add that this finding is consistent over different cluster solutions. Furthermore, all internal cluster validity indices were in agreement with our stability and reproducibility results, indicating that they help finding the optimal cluster solution.

Strengths and limitations

To our knowledge, this is the first study to examine and compare reproducibility of different pattern solutions derived from PCA and cluster analysis. This study is further strengthened by the large study population consisting of both male and female participants, but is limited by the self-reported dietary intake, although this was measured through validated FFQs (13, 14). Moreover, although quantitative criteria can remove some subjectivity involved in deciding on the number of patterns to retain, they often depend on given cutoff values, e.g., the cutoffs of 0.25 for PCA and 0.20 for CFA. In addition, other subjective decisions, such as which food groups to include in the analysis, were not addressed here. Furthermore, finding "the optimal cluster solution" might also depend on the clustering method applied (25). In this study, we mainly focused on KCA and PCA, because these methods are most commonly used in dietary pattern analysis. However, EFA may theoretically be more applicable than PCA, because it takes observed variation into account. Therefore, we repeated our analysis using EFA, which provided similar results.

Overall conclusion

Reliability of dietary patterns differed considerably over different solutions from PCA. In contrast, cluster analysis derived generally stable, reproducible clusters across different solutions. Applied quantitative criteria were valuable for cluster analysis but not for PCA. Associations with disease risk are influenced by the number of patterns that are retained, especially when using PCA. Therefore, studies on the associations between dietary patterns and disease risk should report the reasons to choose the retained numbers of patterns. For PCA, split-half techniques with CFA and visual inspection of the scree plot can be used to substantiate this choice. For cluster analysis, internal cluster validity indices can be applied to help determine the number of clusters to retain.

REFERENCES

1. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM, Guenther PM. Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006. J Am Diet Assoc. 2007;107(7):1233-9.

2. De Coster J. Overview of Factor Analysis. 1998; Available from: www.stat-help.com/factor.pdf.

3. Kant AK. Dietary patterns and health outcomes. J Am Diet Assoc. 2004; 104(4):615-35.

4. Ledesma DL, Valero-Mora P. Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. Practical Assessment, Research & Evaluation. 2007; 12(2).

5. Norusis MJ. Cluster Analysis. IBM SPSS Statistics 19 Statistical Procedures Companion. 2011; Pearson 2011.

6. Newby PK, Muller D, Tucker KL. Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. Am J Clin Nutr. 2004;80(3):759-67.

7. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Long-term stability of food patterns identified by use of factor analysis among Swedish women. J Nutr. 2006;136(3):626-33.

8. Quatromoni PA, Copenhafer DL, Demissie S, D'Agostino RB, O'Horo CE, Nam BH, Millen BE. The internal validity of a dietary pattern analysis. The Framingham Nutrition Studies. J Epid Comm Health. 2002; 56(5):381-8.

9. Lau C, Glumer C, Toft U, Tetens I, Carstensen B, Jorgensen T, Borch-Johnsen K. Identification and reproducibility of dietary patterns in a Danish cohort: the Inter99 study. Br J Nutr. 2008;99(5):1089-98.

10. Dekker LH, Boer JM, Stricker MD, Busschers WB, Snijder MB, Nicolaou M, Verschuren WM. Dietary Patterns within a Population Are More Reproducible Than Those of Individuals. J Nutr. 2013;143(11):1728-35.

11. Lo Siou G, Yasui Y, Csizmadi I, McGregor SE, Robson PJ. Exploring statistical approaches to diminish subjectivity of cluster analysis to derive dietary patterns: The Tomorrow Project. Am J Epidemiol. 2011; 173(8):956-67.

12. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, Jansen EH, van Dieren S, Grobbee DE, Peeters PH, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8.

13. Ocke MC, Bueno-de-Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol. 1997;26 Suppl 1:S37-48.

14. Ocke MC, Bueno-de-Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol. 1997;26 Suppl 1:S49-58.

15. Voedingscentrum. Dutch food consumption table 1996.Voedingscentrum Den Haag, the Netherlands.

16. Stricker MD, Onland-Moret NC, Boer JM, van der Schouw YT, Verschuren WM, May AM, Peeters PH, Beulens JW. Dietary patterns derived from principal component- and k-means cluster analysis: long-term association with coronary heart disease and stroke. Nutr Metabol Cardiovasc Dis. 2013; 23(3):250-6.

17. Herings RM, Bakker A, Stricker BH, Nap G. Pharmaco-morbidity linkage: a feasibility study comparing morbidity in two pharmacy based exposure cohorts. J Epi Com Health. 1992;46(2):136-40.

18. Willett WC. Nutritional Epidemiology. New York: Oxford University Press; 1998.

19. Revelle W. psych: Procedures for Psychological, Psychometric and Personality Research. Evaston Illinois, Northwestern University; 2010 [Accessed 24 October 2013]; Available from: http://personality-project.org/r/psych.manual.pdf.

20. Rosseel Y. lavaan: Latent Variable Analysis. 2010 [Accessed 2 March 2011]; Available from: http://cran.r-project.org/web/ packages/ lavaan/index.html.

21. Rosseel Y. lavaan: an R package for structural equation modeling and more. Version 0.4-8 (BETA). 2011 [Accessed 24 May 2011]; Available from: http://users.ugent.be/~yrosseel/lavaan/lavaanIntroduction.pdf.

22. Team RDC. R: A language and environment for statistical computing. Vienna, Austria: 2011.

23. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. Structure of dietary measurement error: results of the OPEN biomarker study. Am J Epi. 2003;158(1):14-21; discussion 22-6.

24. Hennig C. Cluster-wise assessment of cluster stability. Computational Statistics and Data Analysis. 2007(52):258-71.

25. Wang K, Wang B, Peng L. CVAP: Validation for cluster analyses. Data Science J. 2009(8):93.

Supplemental Table 1 Food groups in the European Prospective Investigation Into Cancer and Nutrition NL cohort.

Chicken	Low fiber cereals
Processed meat	Potatoes
Red meat	Boiled vegetables and legumes
Low-fat fish	Raw vegetables
High-fat fish	Fruit
Shellfish	Nuts and seeds; soy products and peanut butter
Eggs	French fries
Fruit juices	Fast food[1]
Wine	Fat and butter
Alcoholic drinks other than wine	Oils and (diet) margarine
Soft drinks with sugar	Cakes and Cookies
Other non-alcoholic drinks	Sugar and sweets
Soups	Low fat dairy products
High fiber bread	High fat dairy products
Low fiber bread	Cheese
High fiber cereals	

[1] Contains spring rolls, Russian salad, pizza and Dutch fried meat snacks

Supplemental Table 2 Mean (±SD) demographic and lifestyle characteristics of the two half-splits in the European Prospective Investigation Into Cancer and Nutrition NL Cohort (*n*= 39,678).

	Derivation sample		Replication sample	
Participants (*n*)	19837		19841	
Age (years)	49.2	± 11.9	49.2	± 11.9
Female %	74.6		74.3	
Systolic BP (mmHg)	126	± 19	126	± 19
Diastolic BP (mmHg)	77.8	± 10.6	77.8	± 10.7
BMI (kg/m^2)	25.7	± 4.0	25.7	± 4.1
WHR	0.82	± 0.08	0.82	± 0.08
Physically active %	34.0		34.1	
Low education %	57.4		57.5	
Total Protein[1] (g/d)	78.2	± 22.3	78.0	± 22.5
Plant Protein[1] (g/d)	28.3	± 9.6	28.3	± 9.7
Animal Protein[1] (g/d)	49.9	± 16.8	49.8	± 16.8
Total Fat[1] (g/d)	81.9	± 29.6	81.5	± 29.5
Saturated Fat[1] (g/d)	32.2	± 12.0	32.0	± 11.9
Monounsat Fat[1] (g/d)	29.2	± 11.4	29.1	± 11.3
Polyunsaturat Fat[1] (g/d)	14.7	± 6.1	14.7	± 6.2
Carbohydrates[1] (g/d)	232	± 76	231	± 76
Mono/Disacharides[1] (g/d)	116	± 45	116	± 45
Polysaccharidesh[1] (g/d)	115	± 44	115	± 45
Fibre[1] (g/d)	23.8	± 6.6	23.7	± 6.7
Total energy (kcal)	2051	± 620	2047	± 623

[1]Energy adjusted with residual method (18). BP, blood pressure. WHR, waist-to-hip ratio.

Supplemental Methods 1 Exploratory factor analysis (EFA)

The scheme of analysis for EFA was comparable to the PCA analysis and is presented in Figure 1. Five factor solutions ranging from 2 to 6 patterns were extracted in the derivation sample. All solutions were subsequently confirmed in the replication sample using CFA (a). Internal reproducibility was checked by performing EFA in the replication sample and the entire study population (b). Finally, the Kaiser criterion, the scree test optimal coordinate and a visual inspection of the scree plot were applied to identify the number of patterns to retain (c). Data analysis was performed with R.2.15.2 (1). EFA with principal axis factoring, based on the correlation matrix and with varimax rotation, was used to extract solutions ranging from 2 to 6 factors in the derivation sample (factor.pa() function in R-package psych) (2). Each of the obtained solutions was then separately confirmed in the replication sample using CFA (cfa(); lavaan) (3, 4). Food groups with factor loadings $\geq|0.25|$ in EFA (FL_{EFA}) were included in the models for CFA. We retained food groups with factor loadings $\geq|0.20|$ in CFA (FL_{CFA}). Patterns that contained only two dominant food groups were excluded for CFA, e.g. pattern D of the 4-factor solution (Supplemental Table 3), as it is recommended to confirm only patterns with at least 3 dominant food group loadings (5). In order to summarize the results, the ratio of food groups not confirmed (n food groups with $FL_{CFA}\leq|0.20|$) to the total number of food groups (n food groups with $FL_{EFA}\geq|0.25|$) and the deviation in factor loadings between EFA and CFA, defined as

$$Deviation(EFA-CFA)=\frac{1}{k}\cdot\sum_{i=1}^{k}100\cdot|(Fl_{EFA}-Fl_{CFA})/Fl_{EFA}|$$

where k= n food groups, were computed. For example, a value of 20% means that FL_{CFA} of a particular food group differs on average 20% from FL_{EFA}. Confirmation success of all factor solutions obtained by EFA is presented in Supplemental Table 3. The two measures for the reliability of the patterns were well in line with each other, i.e. high ratios of food groups not confirmed to the total number of food groups (NC) were mostly accompanied by high deviations in food group loadings between EFA and CFA. Confirmation success differed considerably between patterns within the same solution. Compared to component solutions (PCA) factor solutions (EFA) were generally better confirmed, with mean deviation ranging from 13.3 to 20.3% (compared to 24.5 to 33.0% in PCA). This most likely reflects that CFA is statistically more related to EFA than PCA. Confirmation success differed considerably between different solutions. The 3- and 4-factor solutions were best confirmed, although the latter contained a pattern consisting of only two food groups (4D) (Supplemental Table 4).

After visual inspection of the scree plot, one might select 3 patterns, while the scree test optimal coordinates as well as the Kaiser criterion suggested extracting 8 patterns. EFA and PCA extracted similar underlying patterns, but for PCA component loadings were higher and component solutions contained considerably more food groups. For example, the 3 factor solution (EFA) contained resp. 10, 5 and 8 food groups, while the 3 component solution (PCA) contained resp. 13, 11 and 11 food groups.

Supplemental Table 3 Confirmation success of different factor solutions (exploratory factor analysis) in the European Prospective Investigation Into Cancer and Nutrition NL Cohort (*n*= 39,678).

Pattern Solution	Pattern	Food groups not confirmed/total number of food groups	Deviation[1]	Mean deviation[2]
2	A	0/10	11.8	20.9
	B	3/8	30.0	
3	A	0/10	10.2	14.8
	B	1/5	23.1	
	C	0/8	11.2	
4	A	1/5	22.7	13.3
	B	0/9	10.2	
	C	0/7	7.1	
	D	NA[3]	NA[3]	
5	A	1/6	16.9	20.3
	B	NA[3]	NA[3]	
	C	0/9	18.5	
	D	0/7	10.2	
	E	1/7	35.4	
6	A	0/3	2.2	16.1
	B	NA[3]	NA[3]	
	C	0/7	9.1	
	D	0/7	21.1	
	E	0/7	36.2	
	F	0/5	11.9	

[1] Deviation between food group loadings in EFA and loadings in CFA. EFA, exploratory factor analysis; CFA, confirmatory factor analysis; [2] The mean deviation of a particular pattern solution; [3] This pattern contained only two dominant food groups and was therefore excluded for CFA. NA, not applicable.

EFA in the whole study population led to comparable food group loadings for all patterns as for the split-half samples (data not shown). As for PCA, reliability of dietary patterns differed considerably between different EFA solutions and the scree plot was the only criterion indicating the same solution as the reliability results. EFA and PCA

extracted similar underlying patterns, but component loadings (PCA) were higher than factor loadings (EFA) leading to the inclusion of more food groups in components compared to factors. Theoretically this can be explained by conceptual differences between PCA and EFA (5-8). Since dietary intake estimation contains measurement error (9) and correlations between food group variables are mostly low to moderate, it could be that dietary patterns derived from PCA are inflated.

Supplemental Table 4 Food group composition of factor solutions 2 to 6 from exploratory factor analysis in the European Prospective Investigation Into Cancer and Nutrition NL Cohort (*n*= 39,678).

Factor Solution	Pattern	Positively scoring food groups[1]	Negatively scoring food groups[1]
2	A	French fries, fast food[2], soft drinks with sugar, low-fiber bread, low-fiber cereals	Fruit, low-fat dairy products, high-fiber bread, cakes and cookies, raw vegetables
	B	Shellfish, high-fat fish, raw vegetables, low-fat fish, wine	Potatoes[3], fat & butter[3], sugar and sweets[3]
3	A	French fries, fast food[2], soft drinks with sugar, low-fiber bread, low-fiber cereals	Fruit, low-fat dairy products, high-fiber bread, cakes and cookies, boiled vegetables and legumes
	B	Shellfish, high-fat fish, low-fat fish, raw vegetables	Sugar and sweets[3]
	C	High-fiber cereals, raw vegetables, nuts[4]	Potatoes, red meat, processed meat, fat & butter, low-fiber bread
4	A	Shellfish, low-fat fish, high-fat fish, raw vegetables	Sugar and sweets[3]
	B	French fries, fast food[2], soft drinks with sugar, low-fiber cereals, other alcoholic drinks	Low-fat dairy products, fruit, cakes and cookies, high-fat dairy products
	C	High-fiber cereals, raw vegetables, nuts[4]	Potatoes, red meat, processed meat, fat & butter
	D	High-fiber bread	Low-fiber bread
5	A	Shellfish, high-fat fish, low-fat fish, raw vegetables, wine	Sugar and sweets[3]
	B	High-fiber bread	Low-fiber bread
	C	Low-fat dairy products, fruit, high-fat dairy products, raw vegetables, high-fiber cereals	French-fries, processed meat, other alcoholic drinks, fast-food
	D	Potatoes, red meat, boiled vegetables and legumes, processed meat	Nuts[4], high-fiber cereals, fast-food

Supplemental table 4 (continued)

Factor Solution	Pattern	Positively scoring food groups[1]	Negatively scoring food groups[1]
5	E	Soft-drinks with sugar, French-fries, fast-food, other non alcoholic drinks, low-fiber cereals[3]	Cakes & cookies, cheese
6	A	Shellfish, high-fat fish, low-fat fish	
	B	High-fiber bread	Low-fiber bread
	C	Raw vegetables, nuts[4], high-fiber cereals	Potatoes, red meat, processed meat, fat & butter
	D	Low-fat dairy products, high-fat dairy products, fruit	Other alcoholic drinks, French fries, fast food[2], processed meat
	E	Soft-drinks with sugar, French fries, fast food[2], other non-alcoholic drinks, low-fiber cereals	Cheese, cakes & cookies
	F	Raw vegetables, boiled vegetables and legumes, fruit, chicken	Sugar and sweets

[1] Food groups are sorted descending according to factor loadings. [2] Contains spring rolls, Russian salad, pizza and Dutch fried meat snacks. [3] Not confirmed food groups. [4] Contains nuts, seeds, soy products and peanut butter

Supplemental Methods 2 Principal component analysis

Scree test optimal coordinate

Visual inspection of the scree plot that graphs the eigenvalues to the component/factor number is widely used to determine the number of patterns to retain (10). According to this criterion, the number of components/factors preceding the scree, which is located where the slope of the graph changes drastically, is extracted. This test has, however, been criticized for being subjective (11). We therefore applied the scree test optimal coordinate, a non-graphical, quantitative solution which uses regression models to estimate predicted eigenvalues for each component/factor. Observed eigenvalues are then compared to the estimated ones, i.e. it is verified whether the observed eigenvalue is superior or equal to the estimated one ($eigenvalue_{observed} \geq eigenvalue_{estimated}$). The last positive verification, beginning at the second eigenvalue, is retained to determine the number of patterns to extract (*nScree*; nFactors) (12, 13).

Kaiser criterion

The Kaiser criterion method is also based on the eigenvalues. It suggests that all components that explain more variance than a single food group variable are extracted. As each single food group variable accounts for an eigenvalue of 1, all components with eigenvalues under 1 are dropped.

Supplemental Methods 3 Cluster analysis

Cluster stability

Cluster stability was measured for all cluster solutions in the derivation sample (*clusterboot();*fpc) (14). The data of the food groups was re-sampled using the bootstrapping technique with a 100 re-sampling runs. In each run, the Jaccard similarities of the original clusters to the most similar clusters in the re-sampled data were computed. The similarity between the sets are measured. An Jaccard similarity of 1 indicates that they are identical, while 0 indicates no common elements between the sets. The mean over these similarities gives information about the stability, i.e. values ≤ 0.5 indicate dissolved clusters, values 0.6-0.75 doubtful clusters, values >0.75 show stable, valid clusters and values >0.85 indicate highly stable clusters (14).

Internal cluster validity indices

Internal cluster validity indices measure the goodness of a cluster solution relative to other solutions and help finding the optimal solution. We computed two indices that

were computationally applicable to our large dataset, i.e. the Calinski-Harabasz- (CH-index, *index.G1();* clusterSim) and the Davies-Bouldin-index (DB-index, *index.DB();* clusterSim), which ranked among the 10 indices considered best (15). The CH-index is based on the ratio of the between- and within cluster sum of squares and the cluster solution with the largest CH-index is considered optimal (15). In order to compute the DB-index, individual cluster indices are first computed for all possible pairs of clusters by summing the mean distance of all subjects in a cluster to their cluster center and dividing this value by the distance of the cluster centers. The DB-index is the mean of the maximum individual indices, i.e. the largest value of each cluster is taken and averaged. The solution with the smallest DB-index is considered optimal (15,16).

REFERENCES

Supplemental methods

1. Team RDC. R: A language and environment for statistical computing. Vienna, Austria: 2011.

2. Revelle W. psych: Procedures for Psychological, Psychometric and Personality Research. Evaston Illinois, Northwestern University; 2010 [Accessed 24 October 2013]; Available from: http://personality-project.org/r/psych.manual.pdf.

3. Rosseel Y. lavaan: Latent Variable Analysis. 2010 [Accessed 2 March 2011]; Available from: http://cran.r-project.org/web/packages/lavaan/index.html.

4. Rosseel Y. lavaan: an R package for structural equation modeling and more. Version 0.4-8 (BETA). 2011 [Accessed 24 May 2011]; Available from: http://users.ugent.be/~yrosseel/lavaan/lavaanIntroduction.pdf.

5. De Coster J. Overview of Factor Analysis. 1998; Available from: www.stat-help.com/factor.pdf.

6. Costello AB, Osborne JW. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. Practical Assessment, Research & Evaluation 10. 2005.

7. Gorsuch RL. Exploratory factor analysis: its role in item analysis. Journal of personality assessment. 1997;68(3):532-60.

8. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. Journal of Psychoeducational Assessment. 2011(29):304-21.

9. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. Journal of the National Cancer Institute. 2011; 103(14):1086-92.

10. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM, Guenther PM. Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006. Journal of the American Dietetic Association. 2007; 107(7):1233-9.

11. Ledesma DL, Valero-Mora P. Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. Practical Assessment, Research & Evaluation 12. 2007.

12. Raiche G, Riopel M, Blais J-G. Non graphical solutions for the Cattell's scree test. 2006 [Accessed 4 October 2011]; Available from: http://www.er.uqam.ca/nobel/r17165/RECHERCHE/COMMUNICATIONS/2006/IMPS/IMPS_PRESENTATION_2006.pdf.

13. Raiche G. nFactors: Parallel analysis and non graphical solutions to the Cattell scree test. 2011 [Accessed October 2011]; Available from: http://cran.r-project.org/web/packages/nFactors/nFactors.pdf.

14. Hennig C. fpc: Flexible procedures for clustering. 2010 [Accessed 24 May 2011]; Available from: http://cran.r-project.org/web/packages/fpc/fpc.pdf.

15. Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a dataset. Psychometrika. 1985;50:159-79.

16. Wang K, Wang B, Peng L. CVAP: Validation for cluster analyses. Data Science Journal. 2009(8):93.

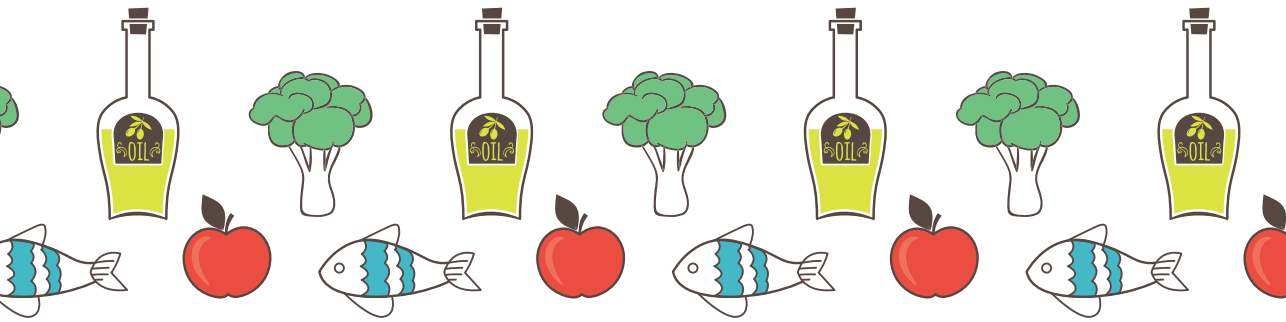
Lifestyle and diet

Chapter 3

Association between lifestyle factors and quality-adjusted life years in the EPIC-NL cohort

Based on: Heidi P Fransen, Anne M May, Joline WJ Beulens, Ellen A Struijk, G Ardine de Wit, Jolanda MA Boer, N Charlotte Onland-Moret, Jeljer Hoekstra, Yvonne T van der Schouw, H Bas Bueno-de-Mesquita, Petra HM Peeters. Association between lifestyle factors and quality-adjusted life years in the EPIC-NL cohort.

PLoS One. 2014 Nov 4;9(11):e111480.
doi:10.1371/journal.pone.0111480



ABSTRACT

The aim of our study was to relate four modifiable lifestyle factors (smoking status, body mass index, physical activity and diet) to health expectancy, using quality-adjusted life years (QALYs) in a prospective cohort study. Data of the prospective EPIC-NL study were used, including 33,066 healthy men and women aged 20-70 years at baseline (1993-7), followed until 31-12-2007 for occurrence of disease and death. Smoking status, body mass index, physical activity and adherence to a Mediterranean-style diet (excluding alcohol) were investigated separately and combined into a healthy lifestyle score, ranging from 0 to 4. QALYs were used as summary measure of healthy life expectancy, combining a person's life expectancy with a weight for quality of life when having a chronic disease. For lifestyle factors analyzed separately the number of years living longer in good health varied from 0.12 year to 0.84 year, after adjusting for covariates. A combination of the four lifestyle factors was positively associated with higher QALYs (P-trend <0.0001). A healthy lifestyle score of 4 compared to a score of 0 was associated with almost a 2 years longer life in good health (1.75 QALYs [95% CI 1.37, 2.14]).

INTRODUCTION

Chronic diseases are important causes of death and disability worldwide. Almost 54% of the global burden of disease is due to non-communicable diseases such as cardiovascular diseases and cancer (1). Nowadays in high income countries, total disease burden is more affected by the years lived with disability than by premature deaths, as the number of premature deaths due to several chronic diseases decreased in the past decades (1, 2). Several known risk factors for these chronic diseases are modifiable, including diet, smoking behavior, weight and physical activity (3). These factors are leading contributors to the global disease burden (2).

The association of modifiable lifestyle factors with either mortality and morbidity has been studied before (4-7). We set out to study the association between a combination of modifiable lifestyle factors and total disease burden using QALYs, a summary health measure that combines life expectancy (mortality) with quality of life (morbidity). QALYs incorporate the effect of different diseases into one outcome measure of health. It provides a more complete picture of the effect of risk factors on population health than using only morbidity and mortality. The QALY was originally developed for use in economic evaluations of health interventions and is generally used as a measure of health gain when comparing interventions (8). One QALY equals one year in optimal health, while zero QALY equals death. Utility weights are applied to represent the reduction in quality of life attributable to having a certain chronic disease. Two previous studies used QALYs to study the association between lifestyle factors and burden of disease (9, 10). One (9) investigated the effect of smoking, physical activity or obesity separately on quality-adjusted life expectancy in the Danish general population. They combined information from life tables with observed age-specific prevalence rates of specific health states to calculate healthy life expectancies (Sullivan's method) (11). Smoking showed the largest impact, followed by physical inactivity and obesity. However, clustering of unhealthy behavior within persons was not taken into account in this study. The other (10) combined lifestyle behaviors and related this to QALYs in a sample of the general population, the EPIC-Norfolk cohort. People with higher health behavior scores had significantly higher QALYs. The study was cross-sectional in that it assessed lifestyle behaviors at baseline in relation to baseline QALY weights. Neither of these studies prospectively followed participants for disease occurrence. The aim of the present study is to relate four modifiable lifestyle factors, smoking status, BMI, physical activity, and adherence to a Mediterranean-style diet, to overall healthy life expectancy, using QALYs. In addition, clustering of unhealthy lifestyle was evaluated by combining lifestyle factors into a healthy lifestyle score. For

this purpose, QALY weights were applied to a large cohort that prospectively followed participants for occurrence of chronic diseases.

METHODS

Study population

The association between lifestyle factors and QALYs was investigated in the EPIC-NL study (12). This prospective cohort study combines the two Dutch cohorts of the European Prospective Investigation into Cancer and Nutrition, i.e. EPIC-MORGEN and EPIC-Prospect (13). 40,011 men and women aged 20-70 years were recruited between 1993 and 1997. At baseline all participants filled out a general questionnaire and a validated food frequency questionnaire, and underwent a physical examination. All participants provided written informed consent before study inclusion. The study complies with the Declaration of Helsinki and was approved by the Institutional Review Board of the University Medical Center Utrecht and the Medical Ethical Committee of TNO Nutrition and Food Research.

Participants were followed for the occurrence of incident diseases by linkage to the Dutch Cancer Registry and the hospital discharge diagnosis database of the National Medical Registry. The National Medical Registry includes information on hospitalized patients; visits to out-patient clinics are not included. Chronic diseases that were identified during follow-up included: diabetes mellitus, stroke, myocardial infarction, other heart conditions, asthma, Chronic Obstructive Pulmonary Disease, osteoarthritis, rheumatoid arthritis, Parkinson's disease and cancer (for this study subdivided into uterus-, bladder-, breast-, colorectal-, skin-, lung-, lymphoma, stomach-, kidney-, prostate-, and other cancer). Diabetes Mellitus was verified using information of the general practitioner or pharmacist (14). Information on date and cause of death was obtained from linkage with municipal registries and Statistics Netherlands. Follow-up information was available until December 31, 2007. Mean follow-up was 12.4 years.

Exclusion criteria were: no written consent for linkage with the registries (n=2879), having one of the investigated chronic diseases at recruitment (this may influence lifestyle at baseline, n=3625), missing dietary information at baseline (n=142), and extreme energy intake levels (being in the top or bottom 0.5% of the ratio of reported energy intake over estimated energy requirement, n=299). Finally, 33,066 participants were included in the analysis.

QALY calculation

The QALY combines information on both life expectancy and quality of life. Years lived with a disease are weighed based on 'utility weights' that vary between 0 and 1. One full year in optimal health equals 1 QALY and death equals 0 QALYs. The utility weights that are used to weigh years lived in a suboptimal health state are in general derived from individual patient health status data, combined with preferences from the general population concerning those specific health states (15). Standardized instruments, such as the EQ-5D instrument (16), have been developed to attach utility values to different health states. EQ-5D health states have been valued by a representative sample of the Dutch population, which resulted in a validated algorithm (15).

To be able to calculate utility weights for specific diseases, information on quality of life is needed from diseased subjects. This information was not available in EPIC-NL. We developed a prediction model for disease-specific utility weights, using data of the second Dutch National Survey of General Practice (DNSGP-2) (17). Details of this prediction model are given in the appendix on page 102. Based on this model utility weights for different diseases, age, sex, education and working status were obtained.

QALYs were estimated for three periods in a person's life: 1. from birth to study recruitment, 2. during the study, and 3. from the end of study (December 2007) until the expected end of life. QALY weights before study entry were based on the participant's gender and age only. Years during the study period were weighted based on disease status and on gender, age, having a paid job and educational level. For each individual all QALYs were summed over the study period. Observed follow-up ended on December 31, 2007 or earlier for death and loss to follow-up. For the calculation of QALYs for the period after study end, we assumed that participants kept the disease state as observed at the end of follow-up until expected date of death. Life expectancy was determined using the participants' age at end of follow-up and the reference year 2007 for national life expectancy statistics. Statistics Netherlands provided age- and gender-specific life expectancy tables (18). For those lost to follow-up before 2007, calendar year of loss to follow up was used as the reference year for life expectancy calculation. For participants who died before 2007 the QALY weight was set to zero from that date onwards.

Lifestyle factors

Smoking status was defined at recruitment as current, past or never smoker. BMI was calculated from measured height and weight and categorized into normal weight (<25 kg/m^2), overweight (25-30 kg/m^2) or obese (≥30 kg/m^2). Physical activity level was assessed in the general questionnaire and categorized according to the validated Cambridge Physical Activity Index (CPAI) in inactive, moderately inactive, moderately active or active (19). Usual food intake as measured by the food frequency questionnaire (20) was used to determine a modified Mediterranean Diet Score (mMDS) (21), with the exception that alcohol consumption was not included in the score (22). However, adjustments for alcohol consumption were made in the analyses. Sex-specific median consumption of eight dietary components was scored for the mMDS, i.e. consumption of fruit, vegetables, legumes, fish, cereals, meat and dairy products and the unsaturated to saturated fat ratio, and summed into a total score that ranged from 0 (minimal adherence) to 8 (maximal adherence). The sum score was categorized into low (0-2), moderate (3-5) and high (6-8) adherence to the mMDS for analysis.

Healthy lifestyle score

In addition to separate analysis of smoking status, BMI, physical activity level and adherence to a Mediterranean-style diet, they were combined into one pragmatic healthy lifestyle score to investigate their combined effect. For the healthy lifestyle score the last three categories of physical activity level, moderately inactive, moderately active and active, were combined into 'being active'. The mMDS was dichotomized into low (0-4) or high (5-8) adherence. Participants scored one point for each of the following lifestyle categories: never smoking, having a normal weight, being physically active and having high adherence to the Mediterranean diet. Hence, the healthy lifestyle score ranged from 0 (unhealthy lifestyle) to 4 (healthy lifestyle). Alcohol consumption was not included as a healthy lifestyle in the score. In most observational studies moderate alcohol intake is associated with decreased risks of cardiovascular disease and diabetes, but any consumption of alcohol is a risk factor for some cancers (23). Furthermore, any alcohol use may increase the risk of binge drinking and alcohol abuse causes health-related harms and higher disease burden (2, 24). Therefore, as we have done previously (22), we decided not to include alcohol consumption in the score, but adjusted for it in the analyses.

Covariates

Age, gender, educational level, alcohol consumption and energy intake were assessed at recruitment. Educational level was categorized into low (lower vocational training or

primary school), middle (secondary school or intermediate vocational training) and higher education (higher vocational training or university). Alcohol intake was included in the model in 7 categories: 0 g/day, ≤6 g/day, 6-≤12 g/day, 12-≤24 g/day, 24-≤60 g/day, >60 g/day (women) or 60-≤96 g/day (men) and more than 96 g/day (men). Energy intake (kcal/day) was used as a continuous variable.

Statistical analysis

Missing data on smoking status, BMI, CPAI, educational level and working status were imputed using single imputation regression modeling (SPSS-MVA). Percentage of missing's ranged from 0.1 to 3% for the different variables. Information on physical activity was missing in 14% of the participants, as in the first year of the EPIC-MORGEN study (1993) physical activity was not assessed with the EPIC questionnaire. Population characteristics according to healthy lifestyle score were presented as mean and standard deviation, median or as a percentage. The association of the separate lifestyle factors or the healthy lifestyle score with QALYs was estimated by linear regression. Regression coefficients and 95% confidence intervals are presented. Additionally, adjustments were made for age at recruitment, gender, educational level, alcohol and energy intake. Analyses for the separate lifestyle factors were additionally adjusted for the other lifestyle factors in the score. All analyses were stratified for cohort. A linear P for trend was computed by including the lifestyle factor or healthy lifestyle score as a continuous variable. Effect modification by gender, age, and educational level was explored by including interaction terms in the model. We performed sensitivity analyses to investigate the effect of excluding participants with a BMI <18 and the effect of including waist circumference instead of BMI in the healthy lifestyle score (1 point if waist is below 94 cm for males or below 80 cm for females). Some diseases that were included in the utility weight model had large confidence intervals around the regression coefficients (see Table A-2 in the appendix). We repeated the analysis and used the lower and upper bounds of the 95% confidence intervals instead of the mean values. Statistical analyses were conducted using SAS 9.2 (SAS Institute, Cary, US).

RESULTS

Table 1 shows characteristics of the study population according to healthy lifestyle score. Participants with the highest score were more often female, young, working and higher educated than participants with a low score.

Table 1 Characteristics of 33,066 healthy EPIC-NL participants according to a healthy lifestyle score.

	Healthy lifestyle score				
	0 n=621 (2%)	1 n=7192 (22%)	2 n=13824 (42%)	3 n=9215 (28%)	4 n=2214 (7%)
Sex (male, %)	37	31	25	23	23
Age at study entry (mean(sd))	53.0(9.4)	50.8(10.3)	49.3(11.7)	47.1(12.7)	45.2(13.3)
Working (yes, %)	36	57	63	68	72
Higher educational level (%)	11	14	18	27	35
Alcohol intake (median(IQR), g/day)	5 (0-23)	6 (1-20)	5 (1-16)	5 (1-14)	4 (1-11)
Total energy intake (mean(sd), kcal/day)	1976(584)	2044(609)	2057(614)	2082(610)	2113(590)
Never smokers (%)	0[a]	5	32	62	100[a]
BMI <25 kg/m^2 (%)	0[a]	6	43	75	100[a]
Physically active (%)	0[a]	85	96	99	100[a]
mMDS score 5-8 (%)	0[a]	4	28	63	100[a]

[a] By definition (a score of 4 implies all group members are never smoker, physically active and have a normal BMI and a high mMDS score)

Table 2 shows results for the individual lifestyle factors. The regression coefficients can be interpreted as the higher (or lower) number of QALYs that an individual has compared to individuals with score zero. Compared to current smokers, never smokers had 0.84 healthy years longer (95% CI: 0.72,0.95), whereas past smokers

showed 0.65 (95% CI: 0.53,0.77) QALYs. Normal weight was associated with 0.44 (95% CI: 0.29,0.59) higher QALYs compared to obesity. Participants who were physically active had 0.73 (95% CI: 0.54,0.92) higher QALYs compared to inactive participants.

Table 2 Regression Coefficients for the relation between separate lifestyle factors and QALYs (N=33,066).

	N	Mean QALY	Crude	P for trend	Adjusted[a]	P for trend
Smoking status						
Current	10035	74.16	reference	<0.0001	reference	<0.0001
Past	10251	75.15	0.99 (0.87, 1.11)		0.65 (0.53, 0.77)	
Never	12780	75.32	1.16 (1.05, 1.28)		0.84 (0.72, 0.95)	
BMI category						
Obese *≥30 kg/m²*	4372	74.64	reference	<0.0001	reference	<0.0001
Overweight *25-30 kg/m²*	13142	74.83	0.19 (0.04, 0.34)		0.25 (0.10, 0.39)	
Normal weight *<25 kg/m²*	15552	75.06	0.43 (0.28, 0.57)		0.44 (0.29, 0.59)	
Physical activity level						
Inactive	2273	74.12	reference	<0.0001	reference	<0.0001
Moderately inactive	8038	74.89	0.76 (0.56, 0.97)		0.48 (0.29, 0.68)	
Moderately active	8696	74.98	0.86 (0.66, 1.06)		0.57 (0.37, 0.77)	
Active	14059	75.02	0.90 (0.70, 1.09)		0.73 (0.54, 0.92)	
mMDS						
0-2	5159	74.63	reference	<0.0001	reference	0.01
3-5	22673	74.93	0.30 (0.17, 0.43)		0.11 (-0.02, 0.24)	
6-8	5234	75.12	0.48 (0.32, 0.65)		0.12 (-0.05, 0.28)	

[a] Adjusted for age at baseline, gender, educational level, alcohol and energy intake and the other lifestyle factors.

Adherence to the mMDS was significantly associated with higher QALYs when added as a continuous variable to the model (p=0.01). Participants who reported to comply with all four lifestyle factors had statistically significantly more QALYs compared to participants that did not comply (1.75; 95% CI: 1.37,2.13), i.e. they lived almost 2 years longer in good health (table 3). The greatest statistically significant association between QALY and lifestyle was observed for a change in healthy lifestyle score from 0 to 1 (0.97; 95% CI: 0.62,1.32), while the association was lower for changes from 3 to 4 (0.17; 95% CI: -0.03,0.37).

Sensitivity analysis, excluding participants with a BMI below 18 (n=187) or including waist circumference instead of BMI in the score, did not alter these results. Also, using the lower and upper bound 95% CI for the EQ-5D utility weight analysis did not affect the results (data not shown). Interactions between healthy lifestyle score and gender, age or educational level were not significant.

Table 3 Regression coefficients for the relation between healthy lifestyle score and QALYs (N=33,066).

	N	mean QALY	Adjusted[a]	P for linear trend
Healthy lifestyle score				
0	621	73.40	reference	<0.0001
1	7192	74.41	0.97 (0.62, 1.32)	
2	13824	74.93	1.35 (1.01, 1.70)	
3	9215	75.25	1.58 (1.23, 1.93)	
4	2214	75.48	1.75 (1.37, 2.13)	

[a] Adjusted for age at baseline, gender, educational level, alcohol and energy intake

DISCUSSION

In this large prospective study, nonsmoking, normal BMI, higher physical activity level and adherence to a Mediterranean-style diet were all associated with a longer healthy life. Being non-smoker and physically active were associated with the highest number of quality-adjusted life years. Furthermore, a combination of the four lifestyle factors was associated with a significantly longer life in good health. People who reported to have never smoked, have a normal weight, are physically active and who adhere to the mMDS lived on average almost two healthy years of life longer compared to people with a less healthy lifestyle.

Strengths of our study are its prospective design, large sample size and the use of a summary health measure to investigate healthy life expectancy. The use of QALYs allowed us to investigate the effect of lifestyle factors on overall health. At the end of the follow-up period (2007), 20% of the cohort suffered from at least one disease and only 4.5% died. Ideally, the cohort should be followed until it is extinct. In the present study, the true beneficial effects of lifestyle factors may therefore have been underestimated. Participants were relatively healthy at study entry, because participants with a prevalent disease at baseline were excluded to rule out the risk of reverse causation. Furthermore, presumably, people with a disease live shorter than the average life expectancy and people without disease live longer than the average life expectancy. Different life expectancies according to disease status were not taken into account, as these data were not available. Moreover, after follow-up was ended, it was not possible to account for new diseases, which probably will develop more frequently in people with an unhealthy lifestyle. That would result in larger differences in QALYs between participants having a healthy or an unhealthy lifestyle. Therefore, our QALYs observed with a healthy lifestyle are likely to be a minimum estimate of the true association.

We could not use our study population to derive the utility weights. Therefore, information on utility weights of the DNSGP-2 study was used (17). As the DNSGP-2 consists of a representative sample of the Dutch GP population and as health-related quality of life was measured through the standardized EQ-5D instrument, we believe that using utility weights obtained from this population is justified. For some diseases, the 95% CI of EQ-5D utility weights were quite large. We therefore investigated the effect of using the lower and upper bound of the 95% CI instead of the mean utility weight in our analysis. This did not alter our results. Regarding endpoint assessment, not all possible lifestyle related diseases could be included. However, we included the most important chronic diseases. Furthermore, several disease data were based on hospital discharge data, while some of the diseases, such as COPD and rheumatoid arthritis, not often require hospitalization. This could have resulted in an underestimation of incidence rates for these diseases. Presuming that these diseases develop more in people with an unhealthy lifestyle, it would result in an underestimation of the QALY difference. Another limitation is that lifestyle factors were only assessed at baseline and possible behavioral changes during follow-up were not taken into account.

Other studies investigated lifestyle factors and health expectancy cross-sectional or applied Sullivan's life table approach using population-based QALYs (9, 10). In our study we computed individual QALYs and adjusted for confounding. Moreover, our study participants were prospectively followed for disease occurrence and their utility

weight was adjusted accordingly, i.e. every time a disease occurred the quality of life was adjusted. For most diseases utility weights are lower in the first year of diagnosis. Assuming that disease utilities stay constant from the start of the disease onwards seems incorrect and may lead to an overestimation of QALYs associated with a healthy lifestyle. Our results are in line with the results of Brønnum-Hansen *et al.*(9) who also reported the greatest benefit of never smoking and being physically active on QALYs. For heavy smoking men almost 10 fewer healthy life years (QALYs) were expected than for never smokers, while we observed only 0.84 fewer QALYs in smokers compared to non-smokers. However, the methods they used differ from our methods: we cannot directly compare the results. They used Sullivan's life table approach together with average age-specific QALY weights to calculate expected QALYs per risk factor exposure level. In Sullivan's life table approach expected QALYs are calculated per risk-factor exposure level and each exposure level results in a different life expectancy. The larger QALY gains found by Brønnum-Hansen *et al.* might be explained by the fact that in the present study the true effect is still underestimated, as discussed above.

Myint *et al.* investigated QALYs and lifestyle behavior (10). They showed that having four healthy behaviors (non-smoking, not physically inactive, drinking 1 to 14 units of alcohol/week and consuming 5 or more servings of fruit and vegetables/week) was related with 1 more healthy year compared to participants with 0 healthy behaviors. In our study almost 2 healthy years were related with adhering to four healthy lifestyle factors. This difference may be explained by the difference in components of the score and the different calculation in QALYs. Myinth *et al.* calculated QALYs until the end of follow-up, while in the present study QALYs were calculated until the end of life by assuming a status quo from end of follow-up to end of life. Finally, Klijs *et al.* studied effects of BMI, smoking and alcohol consumption on years lived with disability and mortality, without using the concept of QALYs. They used the Sullivan life table method (25) and showed that smoking affected mortality more than morbidity, and the other way around for obesity. This supports our method combining mortality and morbidity to investigate total disease burden.

In conclusion, in this prospective study never smoking, having a normal weight, being physically active and adherent to a Mediterranean-style diet were positively associated with healthy life expectancy. The combination of these four lifestyle factors was associated with almost 2 years of life longer in good health. Our findings implicate that public programs aiming at improving health could benefit from targeting at a cluster of modifiable lifestyle factors.

REFERENCES

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2197-223.

2. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369(10):954-64.

3. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2224-60.

4. van den Brandt PA. The impact of a Mediterranean diet and healthy lifestyle on premature mortality in men and women. Am J Clin Nutr. 2011;94(3):913-20.

5. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med. 2012;55(3):163-70.

6. Kvaavik E, Batty GD, Ursin G, Huxley R, Gale CR. Influence of individual and combined health behaviors on total and cause-specific mortality in men and women: the United Kingdom health and lifestyle survey. Arch Intern Med. 2010;170(8):711-8.

7. Knoops KT, de Groot LC, Kromhout D, Perrin AE, Moreiras-Varela O, Menotti A, et al. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. JAMA. 2004;292(12):1433-9.

8. Drummond M, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic Evaluation of Health Care Programmes, 3rd ed. Oxford, UK: Oxford University Press; 2005.

9. Brønnum-Hansen H, Juel K, Davidsen M, Sorensen J. Impact of selected risk factors on quality-adjusted life expectancy in Denmark. Scand J Public Health. 2007; 35(5):510-5.

10. Myint PK, Smith RD, Luben RN, Surtees PG, Wainwright NW, Wareham NJ, et al. Lifestyle behaviours and quality-adjusted life years in middle and older age. Age Ageing. 2011;40(5):589-95.

11. Sullivan DF. A single index of mortality and morbidity. HSMHA Health Rep. 1971;86(4):347-54.

12. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8.

13. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14.

14. Sluijs I, van der AD, Beulens JW, Spijkerman AM, Ros MM, Grobbee DE, et al. Ascertainment and verification of diabetes in the EPIC-NL study. Neth J Med. 2010; 68(1):333-9.

15. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd. 2005;149(28):1574-8.

16. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37(1):53-72.

17. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, van der ZJ. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15(1):59-65.

18. Statistics Netherlands. Levensverwachting; geslacht en leeftijd, vanaf 1950 (per jaar). Accessed May 2012. Available at: http://statline.cbs.nl/.

19. Interact Consortium. Validity of a short questionnaire to assess physical activity in 10 European countries. Eur J Epidemiol. 2012;27(1):15-25.

20. Ocke MC, Bueno-de-Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, et al. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol. 1997;26 Suppl 1:S37-S48.

21. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocke MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ. 2005;330(7498):991.

22. May AM, Romaguera D, Travier N, Ekelund U, Bergmann MM, Kaaks R, et al. Combined impact of lifestyle factors on prospective change in body weight and waist circumference in participants of the EPIC-PANACEA study. PLoS One. 2012; 7(11):e50712.

23. Bagnardi V, Blangiardo M, La Vecchia C, Corrao G. A meta-analysis of alcohol drinking and cancer risk. Br J Cancer. 2001;85(11):1700-5.

24. Room R, Babor T, Rehm J. Alcohol and public health. Lancet. 2005; 365(9458):519-30.

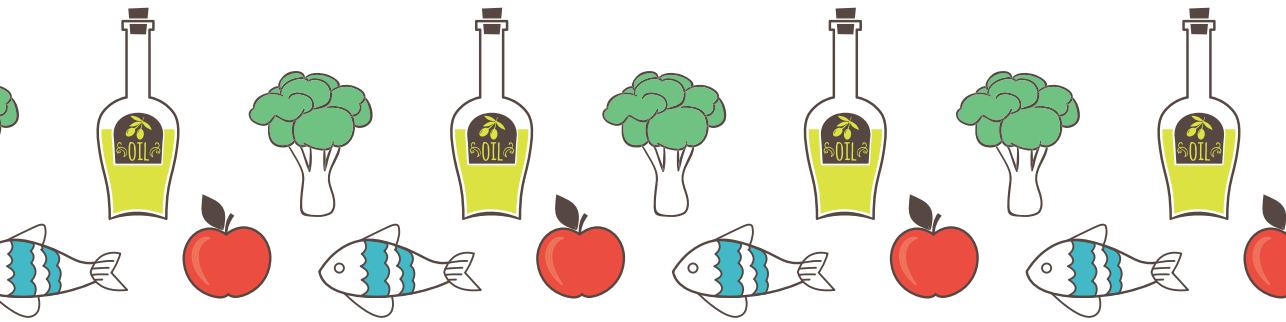
25. Klijs B, Mackenbach JP, Kunst AE. Obesity, smoking, alcohol consumption and years lived with disability: a Sullivan life table approach. BMC Public Health. 2011;11:378.

Chapter 4

Dietary patterns in relation to quality-adjusted life years in the EPIC-NL cohort

Based on: Heidi P Fransen, Joline WJ Beulens, Anne M May, Ellen A Struijk, Jolanda MA Boer, G Ardine de Wit, N Charlotte Onland-Moret, Yvonne T van der Schouw, H Bas Bueno-de-Mesquita, Jeljer Hoekstra, Petra HM Peeters. Dietary patterns in relation to quality-adjusted life years in the EPIC-NL cohort.

Prev Med. 2015 Aug;77:119-24.
doi: 10.1016/j.ypmed.2015.05.014



ABSTRACT

Background: Dietary patterns have been associated with the incidence or mortality of individual non-communicable diseases, but their association with disease burden has received little attention.

Objective: The aim of our study was to relate dietary patterns to health expectancy using quality-adjusted life years (QALYs) as outcome parameter.

Methods: Data from the EPIC-NL study were used, a prospective cohort study of 33,066 healthy men and women aged 20-70 years at recruitment. A lifestyle questionnaire and a validated food frequency questionnaire were administered at study entry (1993-1997). Five dietary patterns were studied: three a priori patterns (the modified Mediterranean Diet Score (mMDS), the WHO-based Healthy Diet Indicator (HDI) and the Dutch Healthy Diet index (DHD-index)) and two a posteriori data-based patterns. QALYs were used as a summary health measure for healthy life expectancy, combining a person's life expectancy with a weight reflecting loss of quality of life associated with having chronic diseases.

Results: The mean QALYs of the participants were 74.9 (standard deviation 4.4). A higher mMDS and HDI were associated with a longer life in good health. Participants who had a high mMDS score (6-9) had 0.17 [95% CI 0.05;0.30] more QALYs than participants with a low score (0-3), equivalent to a two months longer life in good health. Participants with a high HDI score also had more QALYs (0.15 [95% CI 0.03; 0.27]) than participants with a low HDI score.

Conclusion: A Mediterranean-style diet and the Healthy Diet Indicator were associated with approximately 2 months longer life in good health.

INTRODUCTION

Non-communicable diseases such as cancer and cardiovascular diseases are responsible for more than half of the global burden of disease (1). An important modifiable risk factor for many non-communicable diseases is diet (1, 2). In the Netherlands, 5-10% of the burden of disease is related to diet (3). The relation between diet and chronic diseases can be investigated using dietary patterns, where possible correlations and interactions between nutrients or foods are taken into account (4). Dietary patterns can be defined a priori (based on existing knowledge) or a posteriori (based on observed nutritional data). Both a priori- and a posteriori-defined dietary patterns have been used to investigate their relations with the occurrence of chronic diseases (5-10). Adherence to healthier dietary patterns has been associated with lower risk of cardiovascular disease, cancer, and overall mortality in several (5-9), but not all (10) studies.

Studies thus far have mainly focused on associations with the occurrence or mortality of a specific disease. However, diet affects several diseases and even opposing effects may occur. The net effect of diet on disease burden in the population is thus not always clear. We previously used quality-adjusted life years (QALYs) to summarize effects and to account for opposing effects in a prospective cohort (11). This health measure estimates disease burden by including both mortality and morbidity from different diseases while also taking into account severity and duration of the disease. The health status of an individual is reflected in a number between 0 and 1 (utility weight), which represents quality of life associated with disease. QALYs have become popular in cost-effectiveness analysis that compare different health interventions (12). In the present analysis, we computed QALYs in the EPIC-NL cohort and related them to several a priori and a posteriori dietary patterns that are relevant for the Dutch population. The aim of the study was to investigate which dietary pattern is associated with the lowest disease burden.

METHODS

Study population

We investigated the association between dietary pattern adherence and QALYs in the EPIC-NL cohort (13). This cohort consists of EPIC-MORGEN and EPIC-Prospect, the two Dutch cohorts of the European Prospective Investigation into Cancer and Nutrition (EPIC) (14). Between 1993 and 1997, 40,011 men and women aged 20-70 years were recruited. At recruitment, we administered a lifestyle questionnaire and a validated food frequency questionnaire (FFQ) and body weight and height were measured by trained personnel. The lifestyle questionnaire was developed for the EPIC study and contained questions on socio-economic status, occupational history, history of previous illnesses, reproductive history, physical activity and lifetime history of tobacco smoking and alcohol consumption (15). All participants provided informed consent before study inclusion. The study complies with the Declaration of Helsinki and was approved by the Institutional Review Board of the University Medical Center Utrecht and the Medical Ethical Committee of TNO Nutrition and Food Research.

Exclusion criteria for the present analysis were: participants gave no informed consent for linkage with disease registries (n=2879), had one of the investigated chronic diseases at baseline (n=3625), had missing dietary information at baseline (n=142) or unlikely energy intake levels (being in the top or bottom 0.5% of the ratio of reported energy intake over estimated energy requirement, n=299). This resulted in the inclusion of 33,066 participants in the current analysis.

Assessment of dietary intake and dietary patterns

Dietary intake was assessed with a validated FFQ (16) that contained questions on the habitual frequency of consumption of 79 foods during the year preceding enrollment. The FFQ was validated with 12 monthly 24 hour recalls and biomarkers for food groups and nutrients. Median relative validity (Pearson's correlations) was 0.53 for women and 0.61 for men (16, 17). By including information on preparation methods and additions, consumption of 178 food items could be estimated. Food groups (e.g. fruits and vegetables, dairy) were created and nutrient intake was estimated using an extended version of the 1996 computerized Dutch food composition table (18). Food groups and nutrients were combined to construct different dietary patterns.

We investigated three a priori defined dietary patterns. Adherence to a Mediterranean-style diet was assessed by the modified Mediterranean Diet Score (mMDS), developed by Trichopoulou *et al.* (19). The score consists of nine dietary

components: vegetables; legumes; fruit, nuts and seeds; cereals; fish; the ratio of unsaturated to saturated fatty acids; meat; dairy products; and alcohol. Intakes equal to or above the median were assigned a value of 1, and intakes below the median a value of 0. For meat and dairy products, intakes equal or below the median were assigned a value of 1. Alcohol consumption (10-50 g alcohol per day for men or 5-25 g per day for women) was assigned a value of 1. The mMDS score ranges from 0 (minimal adherence) to 9 (maximal adherence). Compared with the original Mediterranean Diet Score (20) fish and poly-unsaturated fatty acids were additionally included in this score.

Furthermore, we studied the Healthy Diet Indicator (HDI), based on WHO recommendations for the prevention of chronic diseases (21). This score includes six nutrients (saturated fatty acids, polyunsaturated fatty acids, cholesterol, protein, dietary fiber, and free sugars) and one food group (fruits and vegetables). Intake within the recommended range was assigned a score of 1. The final HDI score is the sum of all components, ranging from 0 (minimal adherence) to 7 (maximal adherence). The Dutch Healthy Diet index (DHD-index) is based on the Dutch dietary guidelines for a healthy diet, defined by the Dutch Health Council in 2006 (22). This continuous score is based on intake of 8 dietary components: vegetables, fruit, fiber, fish, saturated fatty acids, trans fatty acids, sodium, and alcohol. For all components a maximum score of 10 was assigned when the recommendation was met. Some adjustments were made to the DHD-index for this study. The recommendation to limit consumption occasions to a maximum of 7 was left out because data on food consumption occasions were not available from the FFQ. The physical activity recommendation was also excluded in our score, to increase comparability between the different dietary patterns. No information on added salt was available from our FFQ, but on average its contribution to sodium intake is 30%. Therefore, the cutoff levels for sodium were lowered by 30% to take added salt into account. The final score is the sum of all components, ranging from 0 (minimal adherence) to 80 (maximal adherence). Details of components and scoring of the dietary patterns are presented in Table 1. A higher score represents better adherence to the defined pattern.

In addition to the three a priori defined dietary patterns, we investigated two a posteriori defined dietary patterns. These patterns were obtained through principal component analysis in the EPIC-NL cohort (5) and were labeled 'Western' dietary pattern and 'prudent' dietary pattern.

Table 1 Components of the a priori dietary patterns

		Minimum score per component	Maximum score per component
Modified Mediterranean Diet Score (mMDS), total score 0-9		0 points	1 point
1	Vegetables (excluding potatoes) (g)	< sex-specific median	> sex-specific median
2	Legumes (g)	< sex-specific median	> sex-specific median
3	Fruits, nuts and seeds (g)	< sex-specific median	> sex-specific median
4	Cereals (including bread) (g)	< sex-specific median	> sex-specific median
5	Fish and seafood (g)	< sex-specific median	> sex-specific median
6	Fat ratio (unsaturated	< sex-specific median	> sex-specific median
7	Meat and meat products, incl	> sex-specific median	< sex-specific median
8	Dairy, including cheese (g)	> sex-specific median	< sex-specific median
9	Alcohol (g)	Outside sex-specific range	Sex-specific range: Men: 10-50 g Women: 5-25 g
Healthy Diet Indicator (HDI), total score 0-7		0 points	1 point
1	Saturated fatty acids (en%)	≥10	<10
2	Polyunsat. fatty acids (en%)	<6 or >10	6-10
3	Cholesterol (mg/d)	≥300	<300
4	Protein (en%)	<10 or >15	10-15
5	Dietary fiber (g/d)	<25	≥25
6	Fruits and vegetables (g/d)	<400	≥400
7	Free sugars (en%)	≥10	<10
Dutch Healthy Diet index (DHD-index), total score 0-80		0 points	10 points[a]
1	Vegetables (g/d)	0	≥200
2	Fruit and fruit juices (g/d)[b]	0	≥200
3	Fiber (g/4.2MJ)	0	≥14
4	EPA and DHA (mg/d)	0	≥450
5	Saturated fatty acid (en%)	≥15	<10
6	Trans-fatty acid (en%)	≥1	<1
7	Sodium (g/d)	≥2.52	<1.68
8	Alcohol (g/d)	Men: ≥60 Women: ≥40	Men: ≤20 Women: ≤10
9	Easily fermentable sugars and food acids[c]	> 7 occasions	≤ 7 occasions
10	Physical activity[c]	0 activities	≥5 activities

[a] Each component scores between 0 and 10 on a continuous scale. [b] Maximum of 100 g could be replaced by fruit juice. [c] Components were not included in the DHD-index used in the present study.

The 'Western' pattern represents a high intake of French fries, fast food, low fiber products and soft drinks, while the 'prudent' pattern represents a high intake of fish, vegetables and high-fiber products. Details on these a posteriori patterns can be found in Table 2 and Supplemental table 1.

Table 2 Components of the a posteriori dietary patterns[a].

	Positively scoring food groups[b]	Negatively scoring food groups[b]
Prudent pattern	Raw vegetables, fish and shellfish, wine, fruit, cereals, fruit juices	Potatoes, low-fiber bread, sugar and sweets, fat and butter
Western pattern	French fries, fast food, soft drinks with sugar, low-fiber bread, low-fiber cereals, other alcoholic drinks, processed meat	Fruit, cakes and cookies, high-fiber bread, dairy products, boiled vegetables and legumes, cheese

[a] based on Stricker *et al.*, 2013 (5)
[b] Food group ordering is based on factor loadings, highest to lowest.

Assessment of endpoints

The occurrence of cancer was obtained by linkage to the Dutch Cancer Registry. For this study, cancer was grouped into lymphoma, cancer of the uterus, bladder, breast, colorectal, skin, lung, stomach, kidney, prostate, or other cancer. The occurrence of other chronic diseases (diabetes mellitus, stroke, myocardial infarction, other cardiovascular diseases (such as ischemic heart disease, angina pectoris), asthma, chronic obstructive pulmonary disease, osteoarthritis, rheumatoid arthritis, and Parkinson's disease) was obtained by linkage to the hospital discharge diagnosis database of the National Medical Registry (NMR). The latter database includes information on hospital discharge diagnoses, but visits to outpatient clinics are not included. Diabetes was also ascertained via self-report and a urinary glucose strip test. The general practitioner or pharmacist was contacted to verify the diagnosis of diabetes mellitus (23). Information on date and cause of death was obtained from municipal registries (vital status) and Statistics Netherlands (cause of death). Follow-up information was complete until December 31, 2007.

QALY calculation

In this study, quality adjusted life years were used as the primary outcome for healthy life expectancy. A QALY is calculated by multiplying each year spent in a specific health state by a utility weight that reflects the quality of life for that health state. Utility weights range from 0 to 1, where 1 equals optimal health and 0 equals death. We explained the methods and advantages and disadvantages of computing QALYs in

an ongoing cohort in a previous publication (11). In short, we derived utility weights in data of the second Dutch National Survey of General Practice (DNSGP-2) (24) using the Euroqol health-related quality of life questionnaire (EQ-5D), combined with validated algorithms for the Netherlands (25). A prediction model was used to derive disease-specific utility weights, using sex, age, working status, educational level, and the chronic diseases mentioned previously as predictors (see appendix on page 102). The regression coefficients of this model were applied as utility weights in the EPIC-NL study and for each participant, QALYs were calculated for their entire lifetime. At study entry, participants were free of disease and the accumulated QALYs are based on age at entry, sex, working status, and educational level. Then, annually for each participant, the occurrence of reported diseases was evaluated. If a disease occurred, QALYs were redefined from that time onwards by multiplying the number of years the participant stayed in this disease state by the utility weight that corresponds with the quality of life associated with this disease. We assumed that participants who were still alive at the end of follow-up stayed in the same health state they were in at that time for the rest of their life, i.e., until the end of their life expectancy. Life expectancy was computed using sex-, age- and calendar-specific mortality rates in the Netherlands, provided by Statistics Netherlands (26). Participants' age in 2007 and calendar year 2007 were used as references. For participants who emigrated during follow-up, we used emigration date as the reference year and their age at that date. For participants who died during follow-up, QALY calculation stopped at date of death (utility weight = 0).

Statistical analysis

The association between a dietary pattern (independent variable) and QALYs (outcome) was investigated with linear regression analysis. Regression coefficients and 95% confidence intervals are presented. Analyses were repeated for all 5 dietary patterns. The mMDS and HDI were categorized into low, intermediate and high (0-3, 4-5, 6-9 and 0-2, 3, 4-7, respectively). A low score was used as the reference. The DHD-index and the factor scores for a posteriori patterns were categorized into tertiles. The lowest tertile was used as the reference. Individual component scores of the a priori patterns were associated with QALYs in a sensitivity analysis, using the same model. Associations were adjusted for potential confounders: sex, age, educational level (0.8% missing), physical activity (14% missing), smoking status and intensity (0.4% missing), BMI (0.1% missing), alcohol (3% missing) and energy intake. Educational level was categorized into low (lower vocational training or primary school), middle (secondary school or intermediate vocational training) or high education (higher vocational training or university). Physical activity level was based on hours spent in occupational, cycling and sports activity and categorized into inactive, moderately inactive, moderately

active, or active using the Cambridge Physical Activity Index (27). Smoking status and intensity was categorized into never; former: quit smoking >20 yrs ago, quit 10-20 yrs ago, quit≤ 10 yrs ago; current smoker: <15 cigarettes/day, 15-25 cigarettes/day, >25 cigarettes/day, pipe or cigar smoker. Alcohol intake was categorized into never, quit, 0 g/day, ≤6 g/day, 6-≤12 g/day, 12-≤24 g/day, 24-≤60 g/day, >60 g/day (women) or 60-≤96 g/day and more than 96 g/day (men) and included only in the model if it was not part of the dietary pattern (i.e., HDI). Energy intake (kcal/day) was included as a continuous variable. A P value for linear trend was computed by including the categorical dietary pattern scores as a continuous variable in the model. We studied whether results were comparable for males and females and for smoking status by including interaction terms in the model. Furthermore, to compare the results of the studied dietary patterns we used standardized betas for the scores. A unit change represents a change in 1 standard deviation (SD) in the score. Missing values on confounders were imputed using single linear regression modeling (SPSS). All statistical analyses were conducted using SAS 9.2 (SAS Institute, Cary, US).

RESULTS

Characteristics of the study population are shown in Table 3. Mean expected QALYs of the participants were 74.9 (SD 4.4), which indicates that on average participants are expected to live 74.9 years in good health. More than half (53%) of the population was overweight or obese, almost one-third (30%) were current smokers at recruitment, and 21% had a high educational level.

Components of the five dietary patterns are presented in Table 1 and Table 2. Patterns differ in their included components (food groups and/or nutrients) and used cutoff values. Associations between a priori patterns and QALYs are presented in Table 4. A high modified Mediterranean Diet Score was reported by 24% of the study population, while 41% had a high score on the Healthy Diet Indicator. Participants with an intermediate modified Mediterranean Diet Score (mMDS) had more QALYs than participants with a low mMDS (difference: 0.15 QALYs [95% CI 0.04; 0.26]) (Table 4). Participants with a high mMDS also had more QALYs than participants with a low mMDS (difference: 0.17 QALYs [95% CI 0.05; 0.30]). This is equivalent to living approximately 2 months longer in good health.

Table 3 Characteristics of the EPIC-NL study population (n=33,066).

	%
Male	26
High educational level	21
Current smokers	30
BMI <25 kg/m²	47
Physically active	43
	Mean (sd)
Age at recruitment	48.7 (11.9)
Computed life expectancy	83.5 (5.2)
QALY	74.9 (4.4)
Total energy intake (kcal/day)	2063 (610)
	Median (IQR)
Alcohol intake (g/day)	5.2 (15.2)
Vegetables (g/day)	103.8 (58.3)
Fruit (g/day)	168.1 (158.2)
Fish (g/day)	6.9 (10.2)

Similar results were seen for the Healthy Diet Indicator (HDI); a high or intermediate HDI score was associated with more QALYs than a low HDI score (0.14 [95% CI 0.02; 0.26] and 0.15 [95% CI 0.03; 0.27], respectively). The DHD-index was not significantly associated with QALYs. We investigated which components drive these associations. Components that were individually associated with QALYs were fruits and alcohol for the mMDS and saturated fatty acids, cholesterol, and fruits and vegetables for the HDI (Supplemental table 2). To compare the results of the dietary patterns studied we used standardized betas for the scores. A one SD change in mMDS score was associated with 0.07 QALYs [95% CI 0.03; 0.12], while this was 0.06 [95% CI 0.01; 0.11] for a one SD change in HDI.

In Table 5 the results of the a posteriori dietary patterns are presented. The 'prudent' and the 'western' dietary pattern were not associated with QALYs. Including BMI into the models did not alter the results (data not shown). No interaction was observed between dietary pattern adherence and sex or smoking status. The results were similar for men and women (P for interaction for different dietary patterns

varied between 0.28 and 0.58); and for current, former, and never smokers (P for interaction varied between 0.08 and 0.45) in all analysis.

Table 4 Regression coefficients and 95% CI for the association between a priori dietary patterns and QALYs. EPIC-NL study population, recruited between 1993 and 1997 (N=33,066).

	Category of dietary pattern score			P for trend	Continuous (per SD)
Modified Mediterranean Diet Score					
Score	0-3	4-5	6-9		1.58
N (%)	9769 (30%)	15288 (46%)	8009 (24%)		
QALYs, mean (sd)	74.7 (4.7)	75.0 (4.4)	75.1 (4.1)		74.9 (4.4)
Crude	Reference	0.26 (0.15; 0.37)	0.40 (0.27; 0.53)	<0.0001	0.16 (0.11; 0.21)
Adjusted[a]	Reference	0.15 (0.04; 0.26)	0.17 (0.05; 0.30)	0.01	0.07 (0.03; 0.12)
Healthy Diet Indicator					
Score	0-2	3	4-7		1.15
N (%)	8239 (25%)	11192 (34%)	13635 (41%)		
QALYs, mean (sd)	74.7 (4.8)	74.9 (4.3)	75.0 (4.2)		74.9 (4.4)
Crude	Reference	0.18 (0.05; 0.30)	0.25 (0.13; 0.37)	<0.0001	0.12 (0.08; 0.17)
Adjusted[a,b]	Reference	0.14 (0.02; 0.26)	0.15 (0.03; 0.27)	0.02	0.06 (0.01; 0.11)
Dutch Healthy Diet index					
Tertiles	Tertile 1	Tertile 2	Tertile 3		
N	11022	11022	11022		
DHD-index, mean (sd)	31.9 (4.5)	41.7 (2.5)	55.2 (7.0)		10.96
QALYs, mean (sd)	74.4 (4.5)	74.9 (4.5)	75.4 (4.2)		74.9 (4.4)
Crude	Reference	0.48 (0.36; 0.59)	0.94 (0.83; 1.06)	<0.0001	0.43 (0.39; 0.48)
Adjusted[a]	Reference	-0.08 (-0.20; 0.04)	0.06 (-0.07; 0.19)	0.34	0.04 (-0.01; 0.10)

[a] Adjusted for sex, age at recruitment, smoking status and intensity, educational level, energy intake, and physical activity; [b] Additionally adjusted for alcohol.

Table 5 Association between a posteriori dietary patterns derived from Principal Component Analysis and QALYs, regression coefficients and 95% CI. EPIC-NL study population, recruited between 1993 and 1997 (N=33,066).

	Tertile 1	Tertile 2	Tertile 3	P for trend	Continuous (per SD)
'Prudent' [b]					
N	11022	11022	11022		33066
QALYs mean(sd)	74.5 (4.6)	74.9 (4.4)	75.3 (4.1)		74.9 (4.4)
Crude	Reference	0.35 (0.23; 0.46)	0.78 (0.66; 0.90)	<0.0001	0.32 (0.27; 0.37)
Adjusted[a]	Reference	0.02 (-0.10; 0.13)	0.07 (-0.05; 0.19)	0.26	0.04 (-0.01; 0.08)
'Western' [b]					
QALYs mean(sd)	75.6 (4.3)	74.9 (4.4)	74.2 (4.4)		74.9 (4.4)
Crude	Reference	-0.67 (-0.79; -0.56)	-1.36 (-1.48; -1.25)	<0.0001	-0.58 (-0.63; -0.54)
Adjusted[a]	Reference	-0.15 (-0.27; -0.03)	-0.07 (-0.21; 0.07)	0.30	-0.03 (-0.09; 0.03)

[a] Adjusted for sex, age at recruitment, smoking status and intensity, educational level, energy intake and physical activity; [b] 'Prudent'= a dietary pattern containing a.o. fish, vegetables and high-fiber products; 'Western'= a dietary pattern containing a.o. French fries, fast food, low fiber products and soft drinks (see Table 2 and Supplemental table 1 for details).

DISCUSSION

Our results show that two of the five studied dietary patterns, the modified Mediterranean Diet Score and the Healthy Diet Indicator, were statistically significantly associated with approximately two months longer healthy life.

The association between dietary patterns and single chronic diseases has been investigated before in EPIC-NL (5, 7, 10) and other populations (6, 8, 9, 19). The DHD-index was not associated with CVD risk and cancer (28, 29); a Mediterranean-style diet was inversely associated with all-cause mortality (6, 19, 30). For the HDI an inverse association was seen with all-cause, CVD and cancer mortality (21, 31), but not with breast cancer risk (32). In the EPIC-NL study no association was found between the HDI and cancer incidence (10) or CVD risk (29). These differences across diseases illustrate the need for an approach that summarizes the overall net relation between a dietary pattern and health. In the present study we used such an approach and calculated QALYs in a prospective cohort study. The net effect of the modified Mediterranean Diet

Score (mMDS) and the Healthy Diet Indicator (HDI) was positive, both patterns were significantly associated with healthy life expectancy.

We investigated five different dietary patterns, including both a priori and a posteriori patterns. These patterns differ in their construction, included components (food groups and/or nutrients) and cutoff values used. A posteriori patterns are data driven, while a priori patterns are based on existing knowledge. Regarding the cutoff values, two of the three a priori patterns we investigated use pre-defined cutoff values (the DHD-index and the HDI), while the mMDS uses the median intake of the study population. All these differences may contribute to differences in associations with QALYs. In our study, the predefined patterns, i.e., mMDS and HDI, generally showed stronger associations with QALYs. We additionally investigated which dietary components of the mMDS and the HDI drive the association with QALYs. Fruits and vegetables, alcohol, saturated fatty acids and cholesterol were individually associated with QALYs. Although similar components were included in the DHD-index, none of the components of the DHD-index were individually associated with QALYs. Cutoff points of the mMDS and the HDI are more robust, you either score 1 point if you comply with the recommendation or 0 points if you do not. In the DHD-index scoring of the components is on a continuous scale range from 0 to 10. This may result in less contrast between the groups.

Elsewhere we reported on the association between five dietary patterns and disease burden using DALYs as an outcome measure in the EPIC-NL cohort (33). The results were very similar: higher mMDS and HDI were related to fewer DALYs, which indicates a longer life in good health. DALYs showed slightly stronger associations than QALYs. Although DALYs and QALYs originate from the same concepts, they are not interchangeable (12). Both measures need information on general life expectancy, disease occurrence and mortality, and disease weights (34). However, the weighting of diseases is different. Disability weights for DALYs reflect the function loss due to a disease, based on medical expert values (34), while utility weights for QALYs represent individual quality of life data during a diseased period and are derived from preferences of the general population. This difference in valuation at least partly explains the stronger associations found when using DALYs instead of QALYs.

The strengths of our study are its large sample size and its prospective design. Using a summary health measure allowed us to combine morbidity and severity of different diseases, and mortality into one measure. Several limitations of the study need to be addressed as well. Dietary patterns were based on a single self-reported FFQ administered at recruitment. Possible dietary changes during follow-up were not taken into account. However, in a study in EPIC-Heidelberg dietary intake was reasonably

constant over time (35). Therefore, we also assume a relatively stable consumption in this adult-elderly population. Diseased persons may change to a healthier diet, which may lower their risk of other diseases. This possible change may have diluted our results. Furthermore, the EPIC-NL cohort is relatively young with a median follow-up of 12.4 years. Associations will become stronger when calculated after longer follow-up or in an extinct cohort. To be able to calculate lifetime QALYs we assumed that participants kept the disease state they had at the end of follow-up until the end of their life expectancy. This likely resulted in an underestimation of the association with diet, as people with an unhealthy diet are more likely to develop additional diseases than people with a healthy diet. Additionally, the life expectancies were computed independent of disease status in our study, while they likely differ for different diseases. However, such data were not available. Another limitation of our study is the inclusion of most diseases is based on hospital discharge data only. This again may have resulted in an underestimation of disease incidence, especially for those diseases that cause loss of quality of life but do not always lead to hospitalization, such as musculoskeletal diseases. Furthermore, we were not able to include mental diseases. This overestimation of healthy life years has probably led to an underestimation of the effect of diet, as diseases and premature death are expected to develop more frequently in people that do not adhere to a healthy diet. Disease and mortality rates in EPIC-NL are much lower compared with the general population: a result of the healthy volunteer effect (36). Furthermore, from among these healthy volunteers, we only included participants mean age 49 years that were free of disease at study entry. Our study population is therefore a selective, healthy sample of the general Dutch population. Therefore, associations of a healthy diet with quality adjusted survival may become larger in the general population. In the EPIC-NL study associations were also substantially stronger when healthy diet was combined with other healthy lifestyle habits (11). However, even an on average two months gain in individual healthy life expectancy results in a substantial healthy life gain on a population level.

In conclusion, reporting a healthy dietary pattern was associated with living approximately two more months in good health. A Mediterranean-style diet and the Healthy Diet Indicator were associated with the lowest disease burden. These results can be used to support the nutrition guidelines.

REFERENCES

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2197-223.

2. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369(10):954-64.

3. Van Grinsven H. Balans van de leefomgeving 2014. Planbureau voor de Leefomgeving; 2014; Available from: http://themasites.pbl.nl/balansvandeleefomgeving/2014/voedsel-en-landbouw/ voedsel-en-gezondheid.

4. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol. 2002;13(1):3-9.

5. Stricker MD, Onland-Moret NC, Boer JM, van der Schouw YT, Verschuren WM, May AM, et al. Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke. Nutr Metab Cardiovasc Dis. 2013;23(3):250-6.

6. Sofi F, Abbate R, Gensini GF, Casini A. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. Am J Clin Nutr. 2010; 92(5):1189-96.

7. Waijers PM, Ocke MC, van Rossum CT, Peeters PH, Bamia C, Chloptsios Y, et al. Dietary patterns and survival in older Dutch women. Am J Clin Nutr. 2006;83(5):1170-6.

8. Mente A, de KL, Shannon HS, Anand SS. A systematic review of the evidence supporting a causal link between dietary factors and coronary heart disease. Arch Intern Med. 2009;169(7):659-69.

9. Romaguera D, Vergnaud AC, Peeters PH, van Gils CH, Chan DS, Ferrari P, et al. Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. Am J Clin Nutr. 2012;96(1):150-63.

10. Berentzen NE, Beulens JW, Hoevenaar-Blom MP, Kampman E, Bueno-de-Mesquita HB, Romaguera-Bosch D, et al. Adherence to the WHO's healthy diet indicator and overall cancer risk in the EPIC-NL cohort. PLoS One. 2013;8(8):e70535.

11. Fransen HP, May AM, Beulens JW, Struijk EA, de Wit GA, Boer JM, et al. Association between Lifestyle Factors and Quality-Adjusted Life Years in the EPIC-NL Cohort. PLoS One. 2014;9(11):e111480.

12. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health Policy Plan. 2006;21(5):402-8.

13. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8.

14. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14.

15. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002; 5(6B):1113-24.

16. Ocke MC, Bueno-de-Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, et al. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol. 1997;26 Suppl 1:S37-S48.

17. Ocke MC, Bueno-de-Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol. 1997;26 Suppl 1:S49-S58.

18. Voedingscentrum. Dutch food consumption table 1996.

19. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocke MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ. 2005;330(7498):991.

20. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, et al. Diet and overall survival in elderly people. BMJ. 1995; 311(7018):1457-60.

21. Huijbregts P, Feskens E, Rasanen L, Fidanza F, Nissinen A, Menotti A, et al. Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and The Netherlands: longitudinal cohort study. BMJ. 1997;315(7099):13-7.

22. van Lee L, Geelen A, van HE, de Vries JH, Van't Veer P, Feskens EJ. The Dutch Healthy Diet index (DHD-index): an instrument to measure adherence to the Dutch Guidelines for a Healthy Diet. Nutr J. 2012;11:49.

23. Sluijs I, van der AD, Beulens JW, Spijkerman AM, Ros MM, Grobbee DE, et al. Ascertainment and verification of diabetes in the EPIC-NL study. Neth J Med. 2010; 68(1):333-9.

24. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, van der ZJ. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15(1):59-65.

25. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd. 2005;149(28):1574-8.

26. Statistics Netherlands. Levensverwachting; geslacht en leeftijd, vanaf 1950 (per jaar). Accessed May 2012. Available at: http://statline.cbs.nl/

27. Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr. 2003;6(4):407-13.

28. Struijk EA, May AM, Beulens JW, Fransen HP, de Wit GA, Boer JM, et al. Adherence to the Dutch Guidelines for a Healthy Diet and cancer risk in the European Prospective Investigation into Cancer and Nutrition-Netherlands (EPIC-NL) cohort. Public Health Nutr. 2014;17(11):2546-53.

29. Struijk EA, May AM, Wezenbeek NL, Fransen HP, Soedamah-Muthu SS, Geelen A, et al. Adherence to dietary guidelines and cardiovascular disease risk in the EPIC-NL cohort. Int J Cardiol. 2014;176(2):354-9.

30. Knoops KT, de Groot LC, Kromhout D, Perrin AE, Moreiras-Varela O, Menotti A, et al. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. JAMA. 2004;292(12):1433-9.

31. Knoops KT, Groot de LC, Fidanza F, Alberti-Fidanza A, Kromhout D, van Staveren WA. Comparison of three different dietary scores in relation to 10-year mortality in elderly European subjects: the HALE project. Eur J Clin Nutr. 2006;60(6):746-55.

32. Cade JE, Taylor EF, Burley VJ, Greenwood DC. Does the Mediterranean dietary pattern or the Healthy Diet Index influence the risk of breast cancer in a large British cohort of women? Eur J Clin Nutr. 2011;65(8):920-8.

33. Struijk EA, Beulens JW, May AM, Fransen HP, Boer JM, de Wit GA, et al. Dietary patterns in relation to disease burden expressed in Disability-Adjusted Life Years. Am J Clin Nutr. 2014;100(4):1158-65.

34. Gold MR, Muennig P. Measure-dependent variation in burden of disease estimates: implications for policy. Med Care. 2002;40(3):260-6.

35. Nagel G, Zoller D, Ruf T, Rohrmann S, Linseisen J. Long-term reproducibility of a food-frequency questionnaire and dietary changes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Heidelberg cohort. Br J Nutr. 2007; 98(1):194-200.

36. Struijk EA, May AM, Beulens JW, van Gils CH, Monninkhof EM, van der Schouw YT, et al. Mortality and cancer incidence in the EPIC-NL cohort: impact of the healthy volunteer effect. Eur J Public Health. 2015;25(1):144-9.

Supplemental table 1 PCA component loadings[a]

	'Prudent' pattern	'Western' pattern
Chicken		
Processed meat		,251
Red meat		
Low fat fish	,570	
High fat fish	,585	
Shellfish	,689	
Eggs		
Fruit juices	,263	
Wine	,388	
Other alcoholic drinks		,289
Soft drinks with sugar		,481
Other non-alcoholic drinks		
Soups		
High-fiber bread		-,406
Low-fiber bread	-,344	,375
High-fiber cereals	,277	
Low-fiber cereals	,261	,360
Potatoes	-,391	
Boiled vegetables and legumes		-,333
Raw vegetables	,620	-,219
Fruit	,291	-,510
Nuts seeds soy peanut		
French fries		,637
Fast food[b]		,586
Fat and butter	-,301	
Oils and margarine		
Cakes and cookies		-,413
Sugar and sweets	-,328	
Low fat dairy products		-,516
High fat dairy products		-,312
Cheese		-,275

[a] Only component loadings >|0.20| are presented; Values are expressed as the mean energy contribution to the total energy intake;[b] Contains spring rolls, Russian salad, pizza and Dutch fried meat snacks.

Supplemental table 2 Regression coefficients and 95% CI for the association between individual components of a priori dietary patterns and QALYs. EPIC-NL study population, recruited between 1993 and 1997 (N=33,066) [a].

	Regression coefficients[b] (95% CI)
Components of the modified Mediterranean Diet Score	
Vegetables (excluding potatoes) (g)	0.05 (-0.04; 0.15)
Legumes (g)	-0.09 (-0.18; 0.01)
Fruits, nuts and seeds (g)	0.12 (0.03; 0.22)
Cereals (including bread) (g)	0.09 (-0.02; 0.19)
Fish and seafood (g)	0.03 (-0.06; 0.12)
Fat ratio (unsaturated fat/saturated fat) (g)	0.04 (-0.05; 0.14)
Meat and meat products, including poultry (g)	0.06 (-0.03; 0.16)
Dairy, including cheese (g)	-0.04 (-0.14; 0.05)
Alcohol (g)	0.23 (0.13; 0.32)
Components of the Healthy Diet Indicator[c]	
Saturated fatty acids (en%)	-0.28 (-0.50; -0.05)
Polyunsaturated fatty acids (en%)	0.06 (-0.04; 0.15)
Cholesterol (mg/d)	0.18 (0.03; 0.33)
Protein (en%)	0.04 (-0.06; 0.14)
Dietary fiber (g/d)	0.08 (-0.04; 0.19)
Fruits and vegetables (g/d)	0.28 (0.17; 0.40)
Free sugars (en%)	0.08 (-0.04; 0.21)
Components of the Dutch Healthy Diet Index	
Vegetables (g/d)	0.01 (-0.01; 0.03)
Fruit and fruit juices (g/d)	0.02 (-0.00; 0.04)
Fiber (g/4.2MJ)	0.03 (-0.01; 0.06)
EPA and DHA (mg/d)	0.00 (-0.02; 0.02)
Saturated fatty acid (en%)	0.00 (-0.01; 0.02)
Trans-fatty acid (en%)	-0.00 (-0.01; 0.01)
Sodium (g/d)	0.01 (-0.01; 0.02)
Alcohol (g/d)	0.01 (-0.01; 0.03)

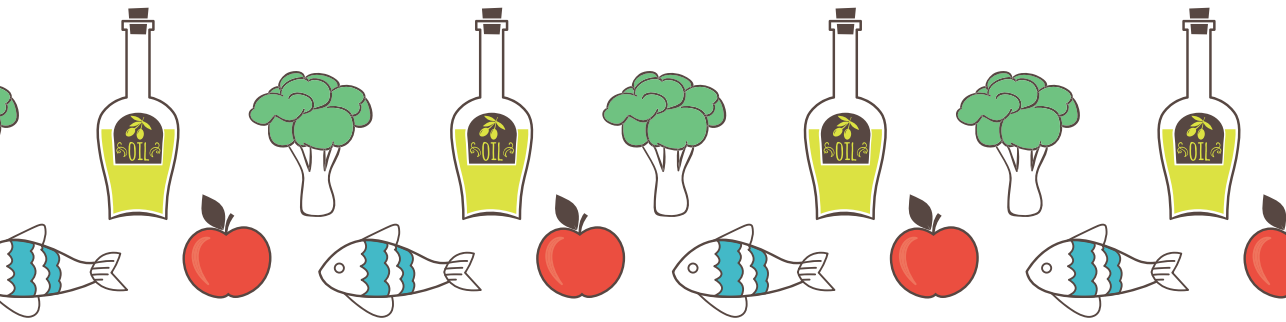
[a] Scores of individual components on the dietary pattern in question are used as an endpoint (see table 1 for scoring details); [b] Adjusted for sex, age at recruitment, smoking status and intensity, educational level, energy intake and physical activity; [c] Additionally adjusted for alcohol.

Chapter 5

Alcohol consumption in relation to chronic disease burden expressed in disability-adjusted life years

Based on: Heidi P Fransen*, Joline WJ Beulens*, Ellen A Struijk, Jolanda MA Boer, G Ardine de Wit, N Charlotte Onland-Moret, Yvonne T van der Schouw, Jeljer Hoekstra, H Bas Bueno-de-Mesquita, Petra HM Peeters, Anne M May. Alcohol consumption in relation to chronic disease burden expressed in disability-adjusted life years. Submitted for publication.

* These authors contributed equally to this paper.



ABSTRACT

Objective: To investigate the association of alcohol consumption with chronic disease burden, calculated as Disability-Adjusted Life Years (DALYs), in a prospective cohort study, thereby accounting for opposing associations between alcohol consumption and different chronic diseases.

Methods: 33 066 healthy men and women aged 20-70 years were recruited into the EPIC-NL cohort during 1993-1997. Alcohol consumption was assessed at baseline with a validated food-frequency questionnaire. Participants were followed until the end of 2007 for occurrence of and mortality from the most important chronic diseases. The association between alcohol consumption categories and DALYs was estimated using a two-part model adjusting for relevant confounders.

Results: After a median follow-up of 12.4 years, 6,647 disease incidences and 1,482 deaths were documented, resulting in 68,225 healthy years of life lost (68,225 DALYs). Moderate drinkers (5-14.9 g for women, 5-29.9 g for men) had a lower chronic disease burden (mean DALYs: -0,27; 95%-CI: -0,43; -0,11), while former drinkers had a higher disease burden (mean DALYs: 0,81; 95%-CI: 0,03; 1,59) than light drinkers (0-4.9 g). Moderate alcohol consumption was associated with a lower disease burden from cardiovascular causes (mean DALYs: -0,18; 95%-CI: -0,29; -0,06) but not from cancer (mean DALYs: -0,05; 95%-CI: -0,16; 0,06). Extreme alcohol consumption was not associated with total DALYs (mean DALYs: -0,02; 95%-CI: -0,28; 0,27). The associations with moderate alcohol consumption were most pronounced among older participants (≥50 years; mean DALYs: -0,32; 95%-CI: -0,53; -0,10), and not visible among younger women (mean DALYs: -0,08; 95%-CI: -0,43; 0,35).

Conclusion: Our results show that moderate alcohol consumption (5-14.9 g/day for women, 5-29.9 g/day for men) was associated with living approximately 3 months longer in good health. These results were mainly observed among middle-aged and older participants and not seen among younger women.

INTRODUCTION

Moderate alcohol consumption, 1-2 drinks per day, has been associated with a reduced risk of all-cause mortality in prospective cohort studies compared to never and heavier drinkers (1). This reduced risk is ascribed to a reduced risk of cardiovascular diseases with moderate alcohol consumption by many researchers (2), while this association is disputed by others (3). Furthermore, any amount of alcohol consumption increases the risk of different forms of cancer such as liver, esophagus, mouth, throat, breast and colorectal cancer (4). In particular, alcohol consumption is linearly associated with approximately a 10% increased risk of breast cancer with each drink per day (4). These opposing associations of alcohol consumption with different chronic diseases emphasizes the importance of investigating the relation of alcohol consumption with total disease burden, which can be estimated with Disability-Adjusted Life Years (DALYs).

To date, DALYs have been mainly calculated on a population level based on statistical data of disease incidence and mortality, like in the Global Burden of Disease Study (GBD) (5). Risk factors such as alcohol consumption are related to estimated DALYs based on effect sizes from observational or intervention studies. According to the Global Burden of Disease Study of 2010, alcohol use accounted for 3.9% of the global disease burden (5). The different results of the cohort studies and GBD can be explained by the fact that the effects of alcohol depend on dosage, which is not accounted for in the GBD studies, and that the net result of alcohol consumption may differ for different populations. Indeed, GBD studies show that the burden of alcohol use was greater among men than women and particularly prevalent in certain regions such as Eastern Europe. These differences are thought to be explained by women drinking less and in less harmful ways than men do and the contribution of injuries and road traffic due to alcohol use to disease burden in certain regions (5-8).

We recently used DALYs at an individual level as an endpoint in a prospective cohort study to study the relation of lifestyle factors with chronic disease burden (9, 10). This approach allows us to account for opposing effects of risk factors on different chronic diseases, for the influence of dosage and to adjust for confounding factors. Using this approach, we aim to investigate the relation of alcohol consumption with chronic disease burden in the EPIC-NL study. In particular, we will address effect modification by gender and age in these analyses.

METHODS

EPIC-NL study

The EPIC-NL study consists of the two Dutch contributions to the European Prospective Investigation into Cancer and Nutrition (EPIC), which were set up simultaneously between 1993 and 1997. The design and rationale of the EPIC-NL study has been described in detail elsewhere (11). In brief, the Prospect-EPIC study includes 17 357 women aged 49 to 70 years living in Utrecht and vicinity who participated in the nationwide Dutch breast cancer screening programme. The MORGEN-EPIC cohort consists of 22 654 men and women aged 20-65 years selected from random samples of the population in three different Dutch towns (Doetinchem, Amsterdam, and Maastricht). At baseline, a general questionnaire and a food-frequency questionnaire (FFQ) were administered and a physical examination was performed. This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the institutional review board of the University Medical Center Utrecht (Prospect) and the Medical Ethical Committee of TNO Nutrition and Food Research (MORGEN). Written informed consent was obtained from all subjects.

From the total cohort (n=40 011) subjects who did not give permission for linkage with disease registries were excluded (n=2879). Furthermore men and women who suffered from any of the studied diseases (Cancer, Coronary Heart Disease (CHD), Cerebrovascular Accident (CVA), Diabetes Mellitus, Chronic Obstructive Pulmonary Disease (COPD), Asthma, Parkinson's disease, Rheumatoid Arthritis, Osteoarthritis, and Inflammatory Bowel Disease (IBD)) at baseline (n=3625) were excluded. Additionally, we excluded subjects without information on dietary intake (n=142) or with implausible high or low scores for total energy intake (those in the top 0.5% and bottom 0.5% of the ratio of reported energy intake over estimated energy requirement based on basal metabolic rate) (n=299). The final study population consisted of 33 066 men and women.

Alcohol consumption

Alcohol consumption was assessed by the general questionnaire and FFQ. The validity of alcohol consumption measured with the FFQ is good as confirmed by a Spearman correlation of 0.87 between the FFQ and 12 24-hour recalls (12, 13). Subjects were asked whether they formerly or currently used alcohol. If doing so currently, they were asked the number of units of alcohol-free beer, beer, white wine, red wine, port/sherry/vermouth and spirits consumed. Subjects indicated their consumption frequency on a daily/weekly/monthly/yearly scale or as never consumed. Alcohol

consumption was determined based on the FFQ by multiplying the alcohol percentage of each beverage by the standard ethanol weight content (5% for beer, 18.5% for fortified wine, 12.5% for red wine, 12% for white wine and 40% for liquor; means from the average sorts of beverages) and was categorized into 5 categories of total alcohol consumption; teetotalers, former drinkers, light drinkers (0-4.9 g/day), moderate drinkers (5-14.9 g/day for women, 5-29.9 g/day for men), and heavier drinkers (≥15 g/day for women and ≥30 g/day for men). Heavier drinkers were further divided into heavy drinkers (15-29.9 g/day for women, 30-59.9 g/day for men) and extreme drinkers (≥30 g/day for women, ≥60 g/day for men) for a sub-analysis. For the separate beverages the amount of alcohol was divided into 5 groups (beer, red wine, white wine, fortified wine and liquor).

Covariates

The general questionnaire included questions on demographics, presence of chronic diseases and risk factors for chronic diseases, such as smoking status and intensity (categorized as never; former: quit smoking >20 y ago, quit 10-20 y ago, quit ≤ 10 y ago; current smoker: 1-15 cigarettes/d, 16-25 cigarettes/d, >25 cigarettes/d; pipe or cigar smoker) and level of education (categorized as very low (only primary school), low (lower vocational education), middle (secondary school or intermediate vocational training) or high (higher vocational training or university)). BMI was calculated from height and weight, which were measured during the physical examination. Usual dietary intakes was obtained from a self-administered validated FFQ containing questions on the usual frequency of consumption of 79 main food items during the year preceding enrolment (12, 13). This questionnaire allowed the estimation of the average daily consumption of 178 foods. Information of the food frequency questionnaire was used to score intake of eight components of a Mediterranean-style diet: vegetables; legumes; fruit, nuts and seeds; cereals; fish; the ratio of unsaturated to saturated fatty acids; meat; and dairy products. For the first six components, intake equal to or above the population median were assigned a value of 1, and intake below the median a value of 0. For meat and dairy products intake equal or below the median were assigned a value of 1. Points were summed into the modified Mediterranean Diet Score, ranging from zero to eight points (14). Physical activity was assessed using the EPIC physical activity questionnaire and categorized according to the validated Cambridge Physical Activity Index (CPAI) (15, 16). This 4-category index (inactive, moderately inactive, moderately active or active) was derived by cross-classifying three questions referring to activities during the last year against classification of work activity. Because there was no information on physical activity (14%), smoking status (0.4%), educational level (0.8%), alcohol intake (categories) (3.3%) and/or BMI (0,1%) for some of the

participants, missing data was imputed using single linear regression modelling (SPSS MVA procedure).

Endpoint assessment

Participants were followed for mortality and morbidity through linkage with several registries. The selection of the diseases was based on their prevalence and disease burden in the Netherlands, but also on the availability of data sources. Information on vital status and the date of death was obtained through linkage with municipal registries. The cause of death was obtained from Statistics Netherlands. Information on disease occurrence (Cancer, CHD, CVA, Diabetes Mellitus, COPD, Asthma, Parkinson's disease, Rheumatoid Arthritis, Osteoarthritis, and IBD) was obtained from the National Cancer Registry and the national hospital discharge diagnosis database from the Dutch National Medical Registry. The National Cancer Registry provided information on the type of cancer and the date of histological diagnosis. The national hospital discharge diagnosis database provided the date of diagnosis for CHD, CVA, Diabetes Mellitus, COPD, Asthma, Parkinson's disease, Rheumatoid Arthritis, Osteoarthritis, and IBD. The national hospital discharge diagnosis database was linked to the cohort with a validated probabilistic method using the following information: date of birth, gender, postal code and code of the general practitioner (17). Self-report and urinary glucose strip tests provided additional information on Diabetes Mellitus at study entry. New Diabetes Mellitus cases ascertained during follow up were verified against information of the general practitioner or pharmacist (18). Follow-up was complete until 31 December 2007.

Computation of DALYs

DALYs, i.e., the sum of the Years Lived with Disability (YLD) and the Years of Life Lost due to premature mortality (YLL) (19), were calculated for each individual in the cohort as previously described (10). The YLL are computed as the number of years death occurred earlier than expected. The expected number of life years are the remaining years that a person of a certain age is expected to live on average defined at time of death, loss of follow-up or end of follow-up (10). The expected life expectancy was obtained from statistics Netherlands that provides age, sex and calendar year specific life expectancies based on mortality rates (20). The YLD are calculated by the number of years a person lives with a disability multiplied by a disability weight reflecting the severity of that disability. The disability weights were derived from the Dutch Disability Weight study, in which 3 panels of medical experts evaluated a large number of

disease stages using techniques such as person trade-off. The disability weights can range between zero (no burden) and one (death) (Supplemental table 1) (21). The years lived with a chronic disease are calculated from the disease onset until death or until the end of life expectancy. One DALY represents the loss of one year in full health. For example, a person who lived five years with diabetes (disability weight 0.20) and died 30 years before his life expectancy obtains one YLD and 30 YLL which equals 31 healthy years of life lost (31 DALYs) (Figure 1). Total DALYs were computed, as well as DALYs for specific diseases: coronary heart disease (CHD), cardiovascular diseases (CVD, includes CHD and cerebrovascular accident), and cancer. DALYs from cancer were further separated into alcohol-related (the upper aero-digestive tract, breast, liver, and colorectal cancer) and non-alcohol related cancer. DALYs for participants that suffered from both alcohol- and non-alcohol-related cancers were included in the alcohol-related cancer DALYs.

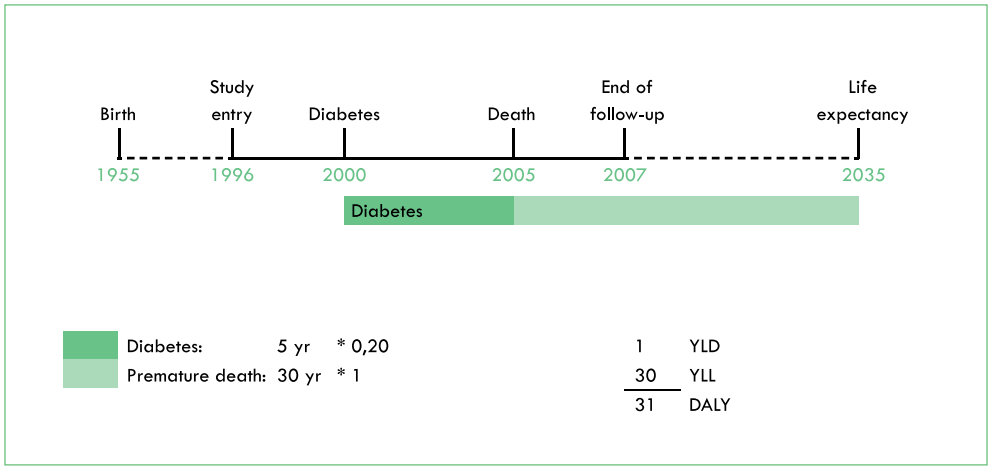


Figure 1 Example of DALY calculation

Statistical analysis

Due to the distribution of the DALYs, i.e. a peak at 0 and a normal distribution in participants with DALYs>0, we used a two-part model to estimate the association of alcohol intake with DALYs as dependent variable (22). This two-part model combines the estimation of the probability of having DALYs using logistic regression with the estimation of the number of DALYs among participants with DALYs > 0 using linear regression. Alcohol intake was analyzed categorically, using light drinkers (0-5 g/d) as the reference. Confidence intervals were constructed with bootstrapping (500 samples).

A quadratic P for trend was investigated by applying a regression with quadratic contrast coefficients to DALY estimates of each of the categories, excluding teetotalers and former drinkers, in all bootstrapping samples. The P value was the percentage of positive (or negative, depending on the direction of the association) regression coefficients of the bootstrap times 2 because of 2 sided testing.

The analyses were adjusted for age, sex, physical activity level (CPAI index), education level, smoking status and intensity and energy intake (kcal/day). In a second model BMI (continuous) and diet (continuous mMDS score) were additionally included as covariates. We investigated interaction with age and sex by including an interaction term in the logistic and linear model. We conducted additional analyses to investigate the effect of stratifying the analysis for sex and age (>50/≤50 years). To exclude confounding by smoking, the analyses were repeated among non-smokers only. Statistical analyses were conducted using SAS 9.2 (SAS Institute, Cary, US), except imputation which has been conducted using SPSS 14.0 (Chicago, IL, USA).

RESULTS

Participants who drank more than 5 grams of alcohol per day were on average older, had a slightly lower BMI, smoked more often and had a higher energy intake than light drinkers (Table 1). After a median follow-up of 12.4 years, 6,647 disease incidences and 1,482 deaths were documented. During the entire follow-up period, 68,225 healthy years of life were lost (68,225 DALYs).

In crude analyses, moderate alcohol consumption was associated with a significantly lower disease burden (mean DALYs: -0,42; 95%-CI: -0,55; -0,28) compared with light alcohol consumption (Table 2). Heavier drinkers (mean DALYs: 0,02; 95%-CI: -0,13; 0,18) and former drinkers (mean DALYs: 0,73; 95%-CI: 0,06; 1,44) had a slightly and significantly higher disease burden, respectively, than light drinkers. Adjusting these analyses for confounders did not substantially alter the results for moderate drinkers and former drinkers: moderate drinkers still had a significantly lower disease burden (mean DALYs: -0,27; 95%-CI: -0,43; -0,11), while former drinkers had a significantly higher disease burden (mean DALYs: 0,81; 95%-CI: 0,03; 1,59) than light drinkers. This corresponds to living approximately 3 months longer in good health for moderate drinkers and approximately 10 months shorter for former drinkers. For teetotalers the association changed from a lower disease burden to a higher disease burden compared with light drinkers after adjustment for confounders. Exchanging BMI for waist circumference in the model did not substantially alter the results. When separating

DALYs from CVD and cancer, we observed that moderate alcohol consumption was only associated with a lower disease burden from cardiovascular causes (mean DALYs: -0,18; 95%-CI: -0,29; -0,06) and not from cancer (mean DALYs: -0,05; 95%-CI: -0,16; 0,06) (Table 2).

When separating heavier drinkers into heavy and extreme drinkers, the association with lower disease burden levelled off at higher levels of intake (Table 3). Extreme alcohol consumption was not associated with total DALYs (mean DALYs: -0,02; 95%-CI: -0,28; 0,27) when adjusted for confounders. Separating these analyses for DALYs from CVD and cancer showed that alcohol consumption remained associated with a lower disease burden from CVD causes (mean DALYs: -0,21; -0,40; -0,01), but with a slightly higher disease burden from cancer (mean DALYs: 0,08; 95%-CI: -0,11; 0,29). Finally, separating DALYs from cancer into alcohol-related and non-alcohol related cancer showed that extreme alcohol consumption was significantly associated with a higher disease burden from alcohol-related cancer (mean DALYs: 0,11; 95%-CI: 0,01; 0,21), but not associated with disease burden from non-alcohol related cancer (mean DALYs: -0,04; 95%-CI: -0,20; 0,14).

Table 1 Characteristics of the EPIC-NL participants at recruitment according to alcohol intake, n=33066

	Alcohol intake categories					
	Teetotaler	Former drinker	Light 0 - <5 g/d	Moderate Women 5 - <15 g/d Men 5 - <30 g/d	Heavy Women 15 - <30 g/d Men 30 - <60 g/d	Extreme Women ≥ 30 g/d Men ≥ 60 g/d
Total, N	2267	363	13831	9983	4589	2033
Women, %	79%	52%	86%	58%	69%	83%
Age at recruitment, years	43±13	46±11	50±12	47±12	50±10	52±8
BMI, kg/m^2	26.1±4.6	25.5±4.2	25.9±4.2	25.3±3.6	25.2±3.5	25.2±3.7
Waist circumference, cm	85±12	87±13	84±11	86±11	85±11	85±11
High education, %	10%	18%	14%	28%	29%	29%
Physically active, %	39%	44%	42%	45%	41%	38%
Current smokers, %	27%	36%	24%	32%	39%	53%
Energy intake, kcal/d	2019± 611	2181± 700	1922± 541	2203± 644	2513± 623	2164± 615
Alcohol intake, g/d *(median IQR)*	0 (0)	0 (0)	1 (2)	10 (7)	24 (12)	41 (26)
mMDS	3.8±1.5	4.1±1.6	3.9±1.5	4.1±1.5	4.1±1.5	4.0±1.5

Table 1 continued

	Alcohol intake categories					
	Teetotaler	Former drinker	Light 0 - <5 g/d	Moderate Women 5 - <15 g/d Men 5 - <30 g/d	Heavy Women 15 - <30 g/d Men 30 - <60 g/d	Extreme Women ≥ 30 g/d Men ≥ 60 g/d
Women, N	1784	187	11888	5747	3187	1691
Age at recruitment, years	43±13	46±12	52±12	50±11	52±9	53±8
BMI, kg/m^2	26.2±4.8	24.9±4.3	25.9±4.3	25.0±3.7	24.9±3.4	25.1±3.7
Waist circumference, cm	84±12	82±11	83±11	81±10	81±9	83±10
High education, %	8%	21%	13%	25%	28%	29%
Physically active, %	37%	41%	42%	44%	41%	37%
Current smokers, %	27%	32%	22%	28%	35%	50%
Energy intake, kcal/d	1906± 530	1844± 480	1828± 459	1906± 449	1916± 441	2002± 450
Alcohol intake, g/d	0 (0)	0 (0)	1 (2)	9 (5)	21 (6)	38 (14)
mMDS	3.8 ± 1.4	4.2 ± 1.6	3.9 ± 1.5	4.1 ± 1.5	4.1 ± 1.5	4.1 ± 1.5
Men, N	483	176	1943	4236	1402	342
Age at recruitment, years	41±12	46±10	42±12	42±11	45±10	46±9
BMI, kg/m^2	25.8±3.6	26.2±4.0	25.6±3.6	25.6±3.3	26±3.4	25.9±3.9
Waist circumference, cm	92±11	94±12	91±11	92±10	94±10	95±11
High education, %	15%	15%	20%	32%	31%	29%
Physically active, %	45%	48%	47%	47%	41%	39%
Current smokers, %	28%	41%	32%	36%	49%	65%
Energy intake, kcal/d	2439± 701	2538 ± 720	2495 ± 642	2606 ± 648	2692 ± 641	2964 ± 695
Alcohol intake, g/d	0 (0)	0 (0)	2 (3)	14 (11)	39 (13)	75 (24)
mMDS	3.9 ± 1.6	4.0 ± 1.6	3.9 ± 1.4	4.0 ± 1.5	4.0 ± 1.5	3.7 ± 1.4

Table 2 Mean (± SE) DALYs in different alcohol consumption categories and regression coefficients (95% CI) of the association between alcohol consumption and DALYs, in 33066 EPIC-NL participants

	Alcohol intake categories					
	Teetotaler	Former drinker	Light	Moderate	Heavier	
			0 - 5 g/d	Women 5 - 15 g/d	Women ≥ 15 g/d	Quadratic
				Men 5 - 30 g/d	Men ≥ 30 g/d	P for trend
N	2267	363	13831	9983	6622	
DALY total						
Mean DALY (SD)	2,0 (5,2)	2,9 (6,9)	2,2 (5,2)	1,8 (4,8)	2,2 (5,4)	
Crude	-0,22 (-0,45; -0,00)	0,73 (0,06; 1,44)	reference	-0,42 (-0,55; -0,28)	0,02 (-0,13; 0,18)	<0,01
Adjusted 1	0,22 (-0,08; 0,52)	0,74 (-0,01; 1,52)	reference	-0,31 (-0,48; -0,15)	-0,22 (-0,40; -0,04)	0,008
Adjusted 2	0,16 (-0,13; 0,46)	0,81 (0,03; 1,59)	reference	-0,27 (-0,43; -0,11)	-0,18 (-0,35; -0,00)	0,024
DALY CHD						
Crude	0,00 (-0,08; 0,09)	0,28 (-0,01; 0,56)	reference	-0,03 (-0,09; 0,02)	-0,01 (-0,08; 0,04)	0,332
Adjusted 1	0,06 (-0,11; 0,23)	0,07 (-0,27; 0,41)	reference	-0,15 (-0,24; -0,06)	-0,18 (-0,28; -0,09)	0,144
Adjusted 2	0,04 (-0,12; 0,22)	0,08 (-0,26; 0,41)	reference	-0,14 (-0,23; -0,05)	-0,17 (-0,27; -0,07)	0,18
DALY CVD (CHD+CVA)						
Crude	-0,02 (-0,14; 0,10)	0,31 (-0,04; 0,69)	reference	-0,09 (-0,16; -0,02)	-0,01 (-0,09; 0,07)	0,012
Adjusted 1	0,08 (-0,14; 0,32)	0,17 (-0,29; 0,66)	reference	-0,19 (-0,30; -0,07)	-0,19 (-0,31; -0,05)	0,076
Adjusted 2	0,06 (-0,15; 0,29)	0,19 (-0,28; 0,66)	reference	-0,18 (-0,29; -0,06)	-0,18 (-0,29; -0,04)	0,100
DALY cancer						
Crude	-0,25 (-0,39; -0,11)	0,13 (-0,30; 0,60)	reference	-0,14 (-0,23; -0,06)	0,07 (-0,04; 0,17)	<0,01
Adjusted 1	-0,09 (-0,28; 0,12)	0,21 (-0,33; 0,78)	reference	-0,06 (-0,17; 0,05)	-0,05 (-0,17; 0,06)	0,508
Adjusted 2	-0,10 (-0,29; 0,11)	0,22 (-0,33; 0,80)	reference	-0,05 (-0,16; 0,06)	-0,04 (-0,17; 0,07)	0,536

Adjusted model 1 is adjusted for age, sex, physical activity, education, energy intake, smoking status and intensity; adjusted model 2= model 1 + BMI + mMDS

Table 3 Regression coefficients (95% CI) of the association between alcohol consumption and DALY in 33066 EPIC-NL participants

	Alcohol intake categories						
	Teetotaler	Former drinker	Light 0 - <5 g/d	Moderate Women 5 - <15 g/d Men 5 - <30 g/d	Heavy Women 15 - <30 Men 30 - <60 g/d	Extreme Women ≥ 30 g/d Men ≥ 60 g/d	Quadratic P for trend
N	2267	363	13831	9983	4589	2033	
DALY total							
Crude	-0,22 (-0,45; -0,00)	0,73 (0,06; 1,44)	reference	-0,42 (-0,55; -0,28)	-0,16 (-0,33; -0,00)	0,44 (0,17; 0,73)	<0,01
Adjusted	0,16 (-0,13; 0,46)	0,81 (0,03; 1,59)	reference	-0,27 (-0,43; -0,11)	-0,26 (-0,45; -0,07)	-0,02 (-0,28; 0,27)	0,004
DALY CHD							
Crude	0,00 (-0,08; 0,09)	0,28 (-0,01; 0,56)	reference	-0,03 (-0,09; 0,02)	-0,00 (-0,07; 0,06)	-0,04 (-0,14; 0,06)	0,948
Adjusted	0,04 (-0,12; 0,22)	0,08 (-0,26; 0,41)	reference	-0,14 (-0,23; -0,05)	-0,15 (-0,25; -0,03)	-0,23 (-0,38; -0,08)	0,624
DALY CVD (CHD+CVA)							
Crude	-0,02 (-0,14; 0,10)	0,31 (-0,04; 0,69)	reference	-0,09 (-0,16; -0,02)	-0,02 (-0,11; 0,08)	0,01 (-0,13; 0,17)	0,20
Adjusted	0,06 (-0,15; 0,29)	0,19 (-0,28; 0,66)	reference	-0,18 (-0,29; -0,06)	-0,16 (-0,30; -0,01)	-0,21 (-0,40; -0,01)	0,36
DALY cancer							
Crude	-0,25 (-0,39; -0,11)	0,13 (-0,30; 0,60)	reference	-0,14 (-0,23; -0,06)	-0,06 (-0,18; 0,04)	0,35 (0,15; 0,55)	<0,01
Adjusted	-0,10 (-0,29; 0,11)	0,22 (-0,33; 0,80)	reference	-0,05 (-0,16; 0,06)	-0,10 (-0,25; 0,01)	0,08 (-0,11; 0,29)	0,032
DALY alcohol-related cancer							
Crude	-0,14 (-0,23; -0,03)	0,02 (-0,26; 0,33)	reference	-0,08 (-0,13; -0,03)	-0,03 (-0,11; 0,03)	0,18 (0,05; 0,31)	<0,01
Adjusted	-0,08 (-0,17; 0,02)	0,10 (-0,19; 0,42)	reference	0,01 (-0,05; 0,06)	0,00 (-0,07; 0,06)	0,11 (0,01; 0,21)	0,104
DALY non-alcohol-related cancer							
Crude	-0,11 (-0,21; 0,01)	0,11 (-0,20; 0,50)	reference	-0,06 (-0,13; 0,01)	-0,03 (-0,12; 0,06)	0,17 (0,02; 0,34)	0,008
Adjusted	0,01 (-0,21; 0,22)	0,09 (-0,34; 0,61)	reference	-0,06 (-0,16; 0,05)	-0,13 (-0,27; -0,02)	-0,04 (-0,20; 0,14)	0,148

Adjusted model is adjusted for age, sex, physical activity, education, energy intake, smoking status and intensity, bmi and mmds; alcohol-related= cancer of the upper aero-digestive tract, breast, liver, and colorectum

When stratifying by age (Table 4), both moderate and heavier alcohol consumption were stronger associated with a lower disease burden among older participants (≥50 years; mean DALYs for moderate drinking: -0,32; 95%-CI: -0,53; -0,10) than among younger participants (<50 years; mean DALYs: -0,14; 95%-CI: -0,38; 0,06), although the interaction between age and alcohol was not statistically significant (p>0.14). Stratifying by sex did not reveal large differences between men and women (p interaction> 0.40). Stratifying by age and sex showed that the lower disease burden with moderate alcohol consumption was present among older women only, but not among younger women.

When we repeated the analyses among non-smokers only, similar results were observed with a lower disease burden for moderate drinkers (mean DALYs: -0.24; 95%-CI: -0.48; -0.04), but no significant associations for other alcohol consumption categories. Separating the different alcoholic beverages showed that moderate consumption of beer and red wine were associated with a lower disease burden, while moderate consumption of the other beverages was not (data not shown).

DISCUSSION

This study investigated the relation of alcohol consumption with total disease burden in the EPIC-NL cohort, thus accounting for opposing effects of alcohol on different diseases and dosage. We found that moderate alcohol consumption (5-14.9 g/day for women, 5-29.9 g/day for men) was associated with a lower disease burden compared to light consumption (0-4.9 g/day). This association was mainly observed among older participants and was, as expected, driven by a lower disease burden due to cardiovascular diseases. The relation was absent among younger women. Former drinkers had a higher disease burden than light consumers, while extreme alcohol consumption was not associated with burden from the selected diseases.

The main strength of our study is that we use a prospective cohort study with a summary health measure as outcome, allowing us to calculate the net association of alcohol consumption with disease burden. In addition, the calculation of DALYs based on individual data of a large prospective cohort enables us to study the direct association of different levels of alcohol consumption with DALYs adjusted for confounders.

However, several limitations need to be addressed. Our estimates probably underestimate the true association due to the left and right truncation of the cohort (10). Participants who were still alive at the end of follow-up were assumed to stay in the same state of health until their expected age of death. In reality, part of the

Table 4 Regression coefficients (95% CI) of the association between alcohol consumption and DALYs in 33066 EPIC-NL participants, stratified by age group and sex

	Alcohol intake categories					
	Teetotaler	Former drinker	Light	Moderate	Heavier	
			0 - 5 g/d	Women 5 - 15 g/d Men 5 - 30 g/d	Women ≥ 15 g/d Men ≥ 30 g/d	Quadratic P for trend
DALY total						
<50 yrs	0,02 (-0,31; 0,35)	1,33 (0,41; 2,36)	reference	-0,14 (-0,38; 0,06)	-0,04 (-0,31; 0,21)	0,216
≥50 yrs	0,19 (-0,23; 0,59)	0,02 (-0,94; 1,09)	reference	-0,32 (-0,53; -0,10)	-0,26 (-0,49; -0,03)	0,10
DALY total						
Men	-0,15 (-0,68; 0,46)	0,78 (-0,19; 1,85)	reference	-0,31 (-0,67; 0,02)	-0,18 (-0,59; 0,18)	0,052
Women	0,22 (-0,10; 0,54)	0,56 (-0,54; 1,72)	reference	-0,25 (-0,42; -0,06)	-0,18 (-0,35; 0,01)	0,064
DALY total						
<50 men	-0,11 (-0,77; 0,59)	1,30 (0,04; 2,62)	reference	-0,26 (-0,68; 0,11)	-0,32 (-0,77; 0,11)	0,556
≥50 men	-0,20 (-1,20; 1,09)	0,12 (-1,38; 1,66)	reference	-0,28 (-0,89; 0,28)	0,15 (-0,55; 0,78)	0,352
<50 women	0,06 (-0,35; 0,54)	1,09 (-0,19; 3,47)	reference	-0,08 (-0,43; 0,35)	0,13 (-0,25; 0,55)	0,352
≥50 women	0,30 (-0,13; 0,73)	0,11 (-1,21; 1,50)	reference	-0,28 (-0,47; -0,05)	-0,28 (-0,51; -0,04)	0,216

Adjusted for age, sex, physical activity, education, energy intake, smoking status and intensity, BMI and mMDS

participants who were still disease-free at the end of follow-up (December 2007) will develop diseases before eventually dying. This is not accounted for in the current analysis. For those living with a disease at the end of follow-up, DALYs may be underestimated as well, since their life expectancy is assumed to be similar to that of a healthy person, while they are more likely to die earlier. Additionally, due to the relatively healthy cohort and exclusion of participants with prevalent diseases at baseline, many participants were still disease-free at the end of follow-up. Underestimation of the association may also be due to our inability to include all diseases, such as Alzheimer's disease, depression, infectious diseases, and certain alcohol-related diseases like liver cirrhosis or accidents. Furthermore, the incidence of some of the diseases that were included is probably underestimated because our data were largely based on hospital discharge diagnoses, which may include only severe cases resulting in hospitalization. Our results, based on an observation time of 12 years in healthy participants, are therefore underestimated. However, due to the non-linear relation the full impact of different levels of alcohol consumption are difficult to extrapolate. Another limitation is the self-reported alcohol intake. However, alcohol consumption from the FFQ correlated well with the twelve 24-hour recalls, suggesting sufficient validity for ranking participants on alcohol consumption. Also, we could not account for drinking pattern and we had relatively few heavy drinkers in our cohort. However, binge drinking does not occur frequently among middle-aged or older persons (>55 years), particularly women (23). It is therefore unlikely that this has influenced the results in this cohort of predominantly middle-aged and older participants. The results for heavier alcohol consumption, however, should be interpreted with caution, particularly because certain alcohol-related diseases were not included in the DALY calculation. Furthermore, we studied relations with disease burden; societal consequences were not included. Finally, we cannot rule out the possibility of residual confounding, as in any observational study.

So far DALYs were mainly used in GBD studies, which aim to define the health status in (different parts of) the world and over time, using population data and modeling. According to the Global Burden of Disease Study of 2010, alcohol use accounted for 3.9% of the global disease burden (5). These estimates are opposite to the reduced disease burden with moderate alcohol consumption observed in our study. GBD studies do not discriminate between different effects of different dosages of alcohol consumption such as accidents and liver cirrhosis, which are more likely due to extreme or binge drinking. Indeed, we observed that the reduced disease burden was only present for moderate drinkers, but not for teetotalers, former or heavier drinkers. Former drinkers show worse health (more DALYs) than teetotalers. It could be that some

former drinkers stopped using alcohol because of health problems either related or not related to alcohol.

This is the first study reporting the associations of alcohol consumption with total chronic disease burden based on real-life data from a prospective cohort. We can therefore not compare our results with previous observational studies as these investigated associations of alcohol consumption with separate diseases. By combining different outcomes in DALYs, we were able to show the net association of alcohol consumption on chronic disease burden. This resulted in a net lower disease burden of approximately 3 months among moderate drinkers compared with light drinkers. This is in line with evidence from cohort studies showing a reduced risk of total mortality with moderate alcohol consumption (1). Two previous studies investigated alcohol consumption in relation to life expectancy in prospective cohort studies (24, 25). The study by Streppel et al. showed that alcohol consumption was associated with a 5-year longer life expectancy at age 50, of which 3 years was attributed to wine consumption (25). Li et al. showed that heavy alcohol consumption (>4 drink/day for men, >1 drink/day for women) was associated with a reduction of 4 years in life expectancy (24). Based on a life table approach, Klijs et al. also estimated that heavy drinking was associated with approximately 2 years lived with disability (26). The effect sizes of these studies were much higher than in our study. This may be due to the fact that the reference categories of both studies included former drinkers at baseline, while we used light drinkers as a reference, but this cannot completely explain the differences. The underestimation of the effect size inherent to using the calculation of DALYs in a non-extinct cohort is likely to account for these differences as well.

In our study, disease burden associated with alcohol consumption differed by age. The reduced chronic disease burden with moderate alcohol consumption was most pronounced among participants aged ≥50 years, and not present among women aged < 50 years. These differences are likely due to different diseases attributing to disease burden of these groups. Since cardiovascular diseases account for the majority of disease burden among older adults, the net effect is mainly determined by the effect of alcohol consumption on cardiovascular disease. It is likely that younger women may suffer more frequently from breast cancer than cardiovascular diseases, resulting in a neutral net effect of alcohol consumption. Indeed, among women <50 years, cancer accounted for over 60% of deaths in our study, while this was lower in the other age/sex groups. In that sense, our results are in line with results from observational studies investigating the association of alcohol consumption with these separate diseases. Alternatively, a selection bias of healthy drinkers surviving until older age could also be involved.

Altogether, our results are in line with previous observational studies showing reduced risks of total mortality with moderate alcohol consumption. They also support the current guidelines allowing moderate alcohol consumption up to 1 drink/day for women and up to 2 drinks/day for men (27, 28). However, our results suggest that these recommendations mainly apply to middle-aged or older populations.

REFERENCES

1. Ferrari P, Licaj I, Muller DC, Kragh Andersen P, Johansson M, Boeing H, et al. Lifetime alcohol use and overall and cause-specific mortality in the European Prospective Investigation into Cancer and nutrition (EPIC) study. BMJ Open. 2014;4(7):e005245.

2. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. BMJ. 2011;342:d671.

3. Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. BMJ. 2014; 349:g4164.

4. Cao Y, Willett WC, Rimm EB, Stampfer MJ, Giovannucci EL. Light to moderate intake of alcohol, drinking patterns, and risk of cancer: results from two prospective US cohort studies. BMJ. 2015;351:h4238.

5. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2224-60.

6. Rehm J, Baliunas D, Borges GL, Graham K, Irving H, Kehoe T, et al. The relation between different dimensions of alcohol consumption and burden of disease: an overview. Addiction. 2010;105(5):817-43.

7. Rehm J, Samokhvalov AV, Neuman MG, Room R, Parry C, Lonnroth K, et al. The association between alcohol use, alcohol use disorders and tuberculosis (TB). A systematic review. BMC Public Health. 2009;9:450.

8. Zaridze D, Maximovitch D, Lazarev A, Igitov V, Boroda A, Boreham J, et al. Alcohol poisoning is a main determinant of recent mortality trends in Russia: evidence from a detailed analysis of mortality statistics and autopsies. Int J Epidemiol. 2009; 38(1):143-53.

9. May AM, Struijk EA, Fransen HP, Onland-Moret NC, de Wit GA, Boer JM, et al. The impact of a healthy lifestyle on Disability-Adjusted Life Years: a prospective cohort study. BMC Med. 2015;13:39.

10. Struijk EA, May AM, Beulens JW, de Wit GA, Boer JM, Onland-Moret NC, et al. Development of methodology for disability-adjusted life years (DALYs) calculation based on real-life data. PLoS One. 2013; 8(9):e74294.

11. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8.

12. Ocke MC, Bueno-de-Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, et al. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol. 1997;26 Suppl 1:S37-S48.

13. Ocke MC, Bueno-de-Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol. 1997;26 Suppl 1:S49-S58.

14. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocke MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ. 2005;330(7498):991.

15. Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr. 2003;6(4):407-13.

16. Interact Consortium. Validity of a short questionnaire to assess physical activity in 10 European countries. Eur J Epidemiol. 2012;27(1):15-25.

17. Herings RM, Bakker A, Stricker BH, Nap G. Pharmaco-morbidity linkage: a feasibility study comparing morbidity in two pharmacy based exposure cohorts. J Epidemiol Community Health. 1992; 46(2):136-40.

18. Sluijs I, van der AD, Beulens JW, Spijkerman AM, Ros MM, Grobbee DE, et al. Ascertainment and verification of diabetes in the EPIC-NL study. Neth J Med. 2010; 68(1):333-9.

19. Murray CJ, Ezzati M, Flaxman AD, Lim S, Lozano R, Michaud C, et al. GBD 2010: design, definitions, and metrics. Lancet. 2012;380(9859):2063-6.

20. Statistics Netherlands. Levensverwachting; geslacht en leeftijd, vanaf 1950 (per jaar). Accessed May 2012. Available at: http://statline.cbs.nl/

21. Melse JM, Essink-Bot ML, Kramers PG, Hoeymans N. A national burden of disease calculation: Dutch disability-adjusted life-years. Dutch Burden of Disease Group. Am J Public Health. 2000;90(8):1241-7.

22. Duan NM, W.G.; Morris, C.N.; Newhouse, J.P. A comparison of alternative models for the demand for medical care. J Bus Econ Stat. 1983;1:115–26.

23. Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula MK, Marks JS. Binge drinking among US adults. JAMA. 2003; 289(1):70-5.

24. Li K, Husing A, Kaaks R. Lifestyle risk factors and residual life expectancy at age 40: a German cohort study. BMC Med. 2014;12:59.

25. Streppel MT, Ocke MC, Boshuizen HC, Kok FJ, Kromhout D. Long-term wine consumption is related to cardiovascular mortality and life expectancy independently of moderate alcohol intake: the Zutphen Study. J Epidemiol Community Health. 2009;63(7):534-40.

26. Klijs B, Mackenbach JP, Kunst AE. Obesity, smoking, alcohol consumption and years lived with disability: a Sullivan life table approach. BMC Public Health. 2011;11:378.

27. Gezondheidsraad. Richtlijnen Goede Voeding 2006. Den Haag: 2006.

28. WCRF. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington: World Cancer Research Fund / American Institute of Cancer Research, 2007.

Supplemental table 1 Disability weights for disabilities and different cancer types

Disabilities	Disability weight	Source
Coronary heart disease	0.29	Dutch disability weight
Stroke	0.61	Dutch disability weight
Diabetes mellitus	0.20	Dutch disability weight
Chronic Obstructive Pulmonary disease	0.31	Dutch disability weight
Asthma	0.08	Dutch disability weight
Parkinson´s disease	0.68	Dutch disability weight
Rheumatoid arthritis	0.53	Dutch disability weight
Osteoarthritis	0.19	Dutch disability weight
Inflammatory bowel disease	0.20	Dutch disability weight
Cancer types		
Digestive		
Stomach	0.59	Dutch disability weight
Oesophagus	0.53	Dutch disability weight
Colon	0.30	Dutch disability weight
Pancreas	0.53	Based on Dutch disability weight for oesophagus cancer
Unspecified parts of billary tract	0.53	Based on Dutch disability weight for oesophagus cancer
Galbladder	0.53	Based on Dutch disability weight for oesophagus cancer
Liver and intrahepatic bile ducts	0.53	Based on Dutch disability weight for oesophagus cancer
Rectum	0.30	Based on Dutch disability weight for colon cancer
Rectosigmoid junction	0.30	Based on Dutch disability weight for colon cancer
Small intestine	0.30	Based on Dutch disability weight for colon cancer
Anus and anal canal	0.30	Based on Dutch disability weight for colon cancer
Other and ill-defined digestive organs	0.30	Based on Dutch disability weight for colon cancer
Respiratory, intrathoracic		
Bronchus and lung	0.54	Dutch disability weight
Heart, mediastinum and pleura	0.54	Based on Dutch disability weight for lung cancer
Accessory sinuses	0.54	Based on Dutch disability weight for lung cancer

Supplemental table 1 (continued)

Disabilities	Disability weight	Source
Larynx	0.54	Based on Dutch disability weight for lung cancer
Male genital organs		
Prostate	0.26	Dutch disability weight
Breast		
Breast	0.26	Dutch disability weight
Hematopoetic and reticuloendothelial		
Hematopoetic and reticuloendothelial	0.24	Dutch disability weight
Female genital organs		
Cervix uteri	0.12	Dutch disability weight
Vulva	0.12	Based on Dutch disability weight for cervix uteri cancer
Ovary	0.30	Based on Dutch disability weight for colon cancer
Corpus uteri	0.12	Based on Dutch disability weight for cervix uteri cancer
Unspecified female genital organs	0.30	Based on Dutch disability weight for colon cancer
Skin		
Skin	0.08	Dutch disability weight
Brain, central nervous system		
Brain	0.54	Based on Australian disability weight for brain cancer
Spinal cord, cranial nerves, other parts of central nervous system	0.54	Based on Australian disability weight for brain cancer
Urinary tract		
Kidney, except renal pelvis	0.26	Based on Dutch disability weight for prostate cancer
Bladder	0.26	Based on Dutch disability weight for prostate cancer
Ureter	0.26	Based on Dutch disability weight for prostate cancer
Mesothelial, soft tissue		
Retroperitoneum and peritoneum	0.30	Based on Dutch disability weight for colon cancer
Other connective and soft tissue	0.30	Based on Dutch disability weight for colon cancer
Thyroid endocrine gland		
Thyroid gland	0.20	Based on Dutch disability weight for thyroid cancer

Supplemental table 1 (continued)

Disabilities	Disability weight	Source
Lip, oral cavity, pharynx		
Pyriform sinus	0.53	Based on Dutch disability weight for oesophagus cancer
Tonsil	0.53	Based on Dutch disability weight for oesophagus cancer
Base of tongue	0.53	Based on Dutch disability weight for oesophagus cancer
Floor of mouth	0.53	Based on Dutch disability weight for oesophagus cancer
Oropharynx	0.53	Based on Dutch disability weight for oesophagus cancer
Nasopharynx	0.53	Based on Dutch disability weight for oesophagus cancer
Gum	0.53	Based on Dutch disability weight for oesophagus cancer
Other unspecified parts of tongue	0.53	Based on Dutch disability weight for oesophagus cancer
Other and ill-defined sites in lip, oral cavity and pharynx	0.53	Based on Dutch disability weight for oesophagus cancer
Bone and articular cartilage		
Bone and articular cartilage of limbs	0.30	Based on Australian disability weight for bone cancer

APPENDIX: DERIVING UTILITY WEIGHTS

Baseline information on quality of life (SF-36 questionnaire) was available for a sample of the EPIC-NL study, but the number of prevalent diseases at baseline within EPIC-NL was too low to estimate utility values associated with chronic disease. Therefore, to calculate QALYs, we developed a prediction model for utility weights for different diseases, using data of the second Dutch National Survey of General Practice (DNSGP-2).

STUDY POPULATION

The DNSGP-2 provides information on disease incidence in 2001 as well as educational level, working status, age and gender of patients registered at 104 general practitioners in the Netherlands (1). In a 5% random sample of this population (n=12,699) a health interview was taken, including questions on chronic disease and health-related quality of life (EQ-5D questionnaire). Details of the study have been reported previously (1). Table A-1 shows some characteristics of the study population.

Information on self-reported chronic disease prevalence as well as health-related quality of life was available for 9654 patients. For types of cancer that occurred in more than 10 patients in this population (uterus, bladder, breast, colorectal, skin, lung, lymphoma, stomach, kidney and prostate cancer) separate utility indices were derived. Cancers occurring less than 10 times in this population were grouped into 'cancer, other type'. Parkinson's disease was not included in the health interview. We used information from the general practitioner's registration to identify prevalent Parkinson's patients.

Health-related quality of life was measured through the standardized EQ-5D instrument of the Euroqol group (2). The EQ-5D has been validated for use in the Netherlands. A validated algorithm (3) attaching utility weights to each of the 243 health states that can be described with EQ-5D was used. Individual factors, such as educational level, age and gender, are taken into account when estimating someone's health related quality of life. Therefore, we derived a model to predict utility weights, using patient characteristics and several chronic diseases as predictors. Patients with missing data on all five EQ-5D questions or on all self-reported chronic diseases were excluded from the analysis (n=25). Missing values in some of the EQ-5D questions and in the questions on chronic diseases, work and educational level were imputed using multiple imputation (SPSS).

Table A-1 Baseline characteristics of study population DNSGP-2 (N=9654).

	DNSGP-2 (N=9654)
Gender (% males)	45%
Age in years (mean(sd))	48.9(17.0)
Working (%)	50%
Higher educational level (%)	21%

MODEL FOR UTILITY WEIGHTS

A linear model was used to derive prediction equations with utility scores as the dependent variable. Predictors in the model were gender, age, working status (categorized into working or not working), educational level (categorized into low, middle and higher education, based on the classification used in the EPIC-NL study (4)), and the chronic diseases diabetes, stroke, myocardial infarction (MI), other heart conditions, asthma/Chronic Obstructive Pulmonary Disease, osteoarthritis, rheumatoid arthritis, Parkinson's disease and cancer (uterus, bladder, breast, colorectal, skin, lung, lymphoma, stomach, kidney, prostate and other cancer).
For some diseases (cancer, stroke and MI) quality of life during disease progression or after diagnosis changes. In the model based on time of diagnosis these diseases were split into 'within 12 months' or 'more than a year ago'. For other heart conditions only information on the 12 months preceding the interview was available. We assumed that the utility weight for a patient with other heart conditions diagnosed more than a year ago was comparable to the weight of a patient who had a MI more than one year ago.

STATISTICAL ANALYSIS

The utility weight prediction model was built in IBM SPSS Statistics for Windows, Version 20 (IBM Corp., Armonk, NY), using linear regression with the method 'enter'. The regression coefficients of the model and 95% confidence intervals are presented in Table A-2. These utility weights were applied to our study population each time a chronic disease occurred to prospectively calculate QALYs for each participant. For example: the utility weight for one year of a 50 year old man, with a higher educational level, a job, diabetes mellitus and a myocardial infarction event less than 12 months ago (first year) is:
0.873 (constant) + 0.022 (male gender) + 50*0.000154 (age) + 0.039 (work) + 0.024 (educational level) -0.028 (diabetes mellitus) -0.004 (myocardial infarction, first year) = 0.9337.
If this person develops bladder cancer in the next year, his utility weight for that year will be:
0.873 (constant) + 0.022 (male gender) + 51*0.000154 (age) + 0.039 (work) + 0.024 (educational level) -0.028 (diabetes mellitus) -0.026 (myocardial infarction, > 1year) – 0.056 (cancer, first year) = 0.8559.

Table A-2 Mean regression coefficients (95% confidence interval) for EQ-5D utility weights derived from 9,654 participants of the Dutch National Survey of General Practice.

	EQ-5D utility weight	
	mean	**(95% CI)**
Constant	0.873	(0.860, 0.887)
Gender (male)	0.022	(0.016, 0.029)
Age (per year)	0.000154	(-0.000074, 0.000382)
Paid Work	0.039	(0.031, 0.046)
Educational level (middle)	0.020	(0.013, 0.028)
Educational level (high)	0.024	(0.016, 0.032)
Diabetes Mellitus	-0.028	(-0.043, -0.012)
Stroke, > 1 year [a]	-0.069	(-0.091, -0.047)
Stroke, first year [a]	-0.153	(-0.201, -0.105)
Myocardial infarction, > 1 year	-0.026	(-0.044, -0.008)
Myocardial infarction, first year	-0.004	(-0.057, 0.048)
Other heart condition, first year	-0.069	(-0.090, -0.048)
Asthma/COPD	-0.050	(-0.062, -0.038)
Osteoarthritis	-0.095	(-0.105, -0.085)
Rheumatoid arthritis	-0.097	(-0.112, -0.082)
Parkinson's disease	-0.307	(-0.399, -0.214)
Cancers		
Uterus, > 1 year	-0.024	(-0.063, 0.015)
Bladder, > 1 year	0.052	(-0.003, 0.107)
Breast, > 1 year	-0.020	(-0.055, 0.016)
Colorectal, > 1 year	-0.055	(-0.102, -0.009)
Skin, > 1 year	-0.019	(-0.048, 0.010)
Lung, > 1 year	0.012	(-0.085, 0.109)
Lymfnode, > 1 year	0.053	(-0.006, 0.111)
Stomach, > 1 year	-0.047	(-0.097, 0.002)
Kidney, > 1 year	0.066	(0.014, 0.117)
Prostate, > 1 year	-0.009	(-0.079, 0.061)
Other cancer, > 1 year	-0.000245	(-0.040, 0.040)
All Cancer, first year	-0.056	(-0.083, -0.029)

[a] 'first year' = diagnosed within one year preceding the interview; '> 1 year' excludes the patients who were diagnosed within one year preceding the interview

REFERENCES APPENDIX

1. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, et al. (2005) Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health 15:59-65.

2. Brooks R. (1996) EuroQol: the current state of play. Health Policy 37:53-72.

3. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. (2005) [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Nederlands tijdschrift voor geneeskunde 149:1574-8.

4. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, et al. (2010) Cohort profile: the EPIC-NL study. Int J Epidemiol 39:1170-8.

Determinants of an unhealthy diet and lifestyle

Chapter 6

Associations between lifestyle factors and an unhealthy diet: a cross-sectional analysis

Based on: Heidi P Fransen, Jolanda MA Boer, Joline WJ Beulens, G Ardine de Wit, H Bas Bueno-de-Mesquita, Jeljer Hoekstra, Anne M May, Petra HM Peeters. Associations between lifestyle factors and un unhealthy diet: a cross-sectional analysis. Submitted for publication.

ABSTRACT

Background: Unhealthy dietary patterns have been associated with other unhealthy lifestyle factors such as smoking and a low physical activity level. Whether these associations are present both in high and low educated individuals is currently unknown.

Methods: We conducted a cross-sectional study in the EPIC-NL cohort, a prospective cohort in the Netherlands of 39,393 men and women aged 20-70 years at recruitment. A lifestyle questionnaire and a validated food frequency questionnaire were administered at study entry (1993-1997). Lifestyle factors included body mass index, waist circumference, smoking status, physical activity level, dietary supplement use and daily breakfast consumption. Low adherence to a Mediterranean-style diet was defined as un unhealthy pattern. Multivariate logistic regression analyses were performed for the total population and by strata of educational level. Relationships were adjusted for sex, age, prevalent diseases, and mutually for all other factors.

Results: Of the total study population, 30% had an unhealthy dietary pattern: 39% in the lowest educated group and 20% in the highest educated group. An inactive lifestyle, a large waist circumference, no dietary supplement use and skipping breakfast were associated with an unhealthy dietary pattern, both in low as well as in high educated participants. Among low-educated participants, current smokers had a greater odds of an unhealthy diet compared to never smokers: OR 1.42 [95% CI: 1.25; 1.61]. This association was not present in the high-educated group. In contrast, in the high-educated group quitting smoking was associated with a more healthy diet: OR 0.80 [0.70; 0.92].

Conclusions: Most associations between lifestyle factors and unhealthy diet were consistent across educational levels, except for smoking. Only among low educated participants, current smokers reported an unhealthier dietary pattern in comparison to never smokers. These results can be used in the development of targeted health promotion strategies.

BACKGROUND

An unhealthy diet is associated with a higher risk of chronic diseases and mortality (1). Up to 4% of the global disease burden has been related to an unhealthy diet (2), making diet one of the important modifiable lifestyle factors. Many studies have related an (un)healthy diet to socio-demographic or lifestyle factors. A low educational level, a low income or a low occupational position have been associated with an unhealthy diet (3), while a higher educational level or occupational position have been associated with a healthy diet (4). Furthermore, physical inactivity, smoking and a young age were all related to an unhealthy diet (5, 6). Unhealthy behaviors tend to co-exist in persons with a low socio-economic status (7, 8). Whether such unhealthy behaviors also co-exist among high educated people has not been investigated to date. This information can be useful in the development of targeted health promotion strategies.

Next to individual behaviors, neighborhood factors might also be associated with an unhealthy diet. Living in deprived neighborhoods has been related to having more unhealthy lifestyle factors in some studies (9, 10), while others did not find a neighborhood effect on dietary intake (11).

This study aimed to investigate in the Netherlands whether associations between unhealthy lifestyle factors and unhealthy diets differ between people with a low and a high educational level. In addition to smoking, weight and physical activity we also studied supplement use and daily breakfast consumption. Our aim is to find out if these factors relate to diet to the same extent in low and high educated persons. These data may be helpful in designing tailored interventions. We used the modified Mediterranean Diet Score (mMDS) to define healthy or unhealthy dietary patterns, because this pattern was related to a decrease in chronic diseases (12) and to a longer health expectancy in our study cohort, the European Prospective Investigation into Cancer and Nutrition–The Netherlands (EPIC-NL) cohort (13).

METHODS

Study population

We used baseline data on education, nutrition and other lifestyle characteristics of the EPIC-NL cohort in the Netherlands (14). Between 1993 and 1997, 40,011 men and women aged 20-70 years were recruited. At recruitment, we administered a lifestyle questionnaire and a validated food frequency questionnaire (FFQ) and participants underwent a physical examination. All participants provided informed consent. The

study complies with the Declaration of Helsinki and was approved by the Institutional Review Board of the University Medical Center Utrecht and the Medical Ethical Committee of TNO Nutrition and Food Research. Exclusion criteria for the present analysis were: missing dietary information at baseline (n=218) or implausible energy intake levels (being in the top or bottom 0.5% of the ratio of reported energy intake over estimated energy requirement, n=400). In total 39,393 participants were included in the current analysis.

Definition of unhealthy dietary patterns

Dietary intake at recruitment was assessed with a validated FFQ (15) that contained questions on the habitual frequency of consumption of 79 foods during the year preceding enrollment. This FFQ was previously validated with 12 monthly 24 hour recalls for food groups and nutrients, resulting in a median relative validity (Pearson's correlations) of 0.53 for women and 0.61 for men (15, 16). By including information on preparation methods and additions, consumption of 178 food items was estimated. Food groups (e.g. fruits and vegetables) were created and nutrient intake was estimated using an extended version of the 1996 computerized Dutch food composition table (17). Food groups and nutrients were combined to construct the modified Mediterranean Diet Score (mMDS) (18). In this score individual intake of food groups and nutrients is compared to the sex-specific median value of the study population. For vegetables, legumes, fruits, nuts and seeds, cereals, fish, and the ratio of unsaturated to saturated fatty acids, intakes equal to or above the median were assigned a value of 1, and intakes below the median a value of 0. For meat and dairy products, intakes equal or below the median were assigned a value of 1. We included alcohol intake in the score and assigned a value of 1 to consumption of 10-50 g (men) or 5-25 g (women) alcohol per day. The mMDS score ranges from 0 (minimal adherence) to 9 (maximal adherence). A higher score represents better adherence to the defined pattern. In this study a low adherence (score 0-3) was classified as having an unhealthy diet. We also computed results excluding alcohol from the score. The mMDS score excluding alcohol ranges from 0 (minimal adherence) to 8 (maximal adherence), a low adherence (score 0-3) was classified as having an unhealthy diet.

Educational level

The lifestyle questionnaire provided data on educational level, based on the highest level of completed school education. For this study we defined low (primary school, N=6441), mid-low (lower vocational training, N=16316), mid-high (secondary school or

intermediate vocational training, N=8590) and high education (higher vocational training or university, N=8046).

Individual factors

The lifestyle questionnaire included information on smoking habits, physical activity, dietary supplement use, and pre-existing diseases (heart disease, diabetes mellitus, cancer). Smoking status was categorized into never, former or current smoker. Physical activity level was based on hours spent in occupational activity, cycling and sports activity and categorized into inactive, moderately inactive, moderately active or active using the Cambridge Physical Activity Index (19). Ever use of dietary supplements was included as a dichotomous variable. Pre-existing diseases were combined into one dichotomous measure for prevalent disease (present or not). Body Mass Index (BMI) and waist circumference were calculated from anthropometric measurements taken during the physical examination. BMI was categorized into normal weight (<25 kg/m^2), overweight (25-30 kg/m^2) and obese (>30 kg/m^2). Waist circumference was categorized based on guidelines of the World Health Organization (20) into normal (females <80 cm, males <94 cm), moderately increased (females 80-88 cm, males 94-102 cm) or large (females ≥88 cm, males ≥102 cm). The food frequency questionnaire included information on breakfast frequency (daily or less). Working status was categorized into having a paid job and being without a paid job. The latter included retired participants.

Neighborhood factors

We used two factors containing information on the neighborhood participants lived in: urbanization level and neighborhood social status score. Statistics Netherlands provided data on urbanization level based on postal code in 1995, corresponding to the EPIC-NL recruitment period (21). Postal code at recruitment was available for 38,086 participants. Urbanization level was categorized as 'non urbanized' (0-499 addresses/km^2), 'hardly urbanized' (500-999 addresses/km^2), 'moderately urbanized' (1000-1499 addresses/km^2), 'strongly urbanized' (1500-2499 addresses/km^2) and 'extremely urbanized' (>2499 addresses/km^2). The Netherlands Institute for Social Research provided us with neighborhood social status scores based on the participant's postal code, in the year 1998. The social status scores represent the social status of a neighborhood, based on: mean income, and the percentage of citizens that were unemployed, had a low income and had a low educational level (22). The score ranged from -4.87 to 1.61, with a higher score indicating a higher neighborhood social status.

Negative scores appear because results are based on factor analysis with the value '0' as the mean score over the years 1998-2010.

Statistical analysis

Characteristics of the study population (N=39,393) are presented as a percentage, a mean with standard deviation or as a median with interquartile range. They are presented for the total study population and separately for participants with a low (N=6441) and participants with a high (N=8046) educational level. We focus on participants with a low or high educational level, because we are interested in the extremes. The association between the individual level factors (independent variables) and low adherence to a Mediterranean-style diet (outcome) was investigated with logistic regression analysis. Odds ratios and 95% confidence intervals are presented. Relationships were adjusted for sex, age, prevalent diseases, and mutually for all other individual factors. Analyses with urbanization level and neighborhood social status score were performed using multilevel logistic modelling (SAS procedure 'glimmix') to take the clustered structure of the data into account (i.e., individuals within neighborhoods). Participants with missing data on dietary supplement use (N=251), urbanization level (N=429) or neighborhood social status score (N=51) were excluded in the corresponding analyses. All multilevel models included a random intercept based on postal code. Interaction between factors and educational level was tested by including interaction terms in the model, including the lowest and highest educational level only (N=14,388). Additionally, we studied above relations by strata of educational level. A p-value for trend was computed by including the lifestyle factor as a continuous variable in the model. We performed sensitivity analyses to investigate the effect of excluding participants with a low BMI (<18.5 kg/m^2). All statistical analyses were conducted using SAS 9.2 (SAS Institute, Cary, US).

RESULTS

Characteristics of the study population are presented in Table 1 for the total population and for the low and high educated group. The majority of our study population (74%) was female. Mean age of the study population was 49.2 years. Primary school (low educational level) was the only education for 16% of the study population, while 20% had a high educational level. A low modified Mediterranean Diet Score (mMDS) was reported by 30% of the total study population, by 39% in the low educated group and by 20% in the high educated group. Participants with a low educational level were older, more often female and more likely to be overweight than participants with a high educational level. Moreover, the low educated group included fewer never smokers and was more often physically inactive. Mean social status score of their neighborhood was lower compared to participants with a high educational level (-0.47 vs -0.28).

Associations between individual and neighborhood factors and an unhealthy diet are presented in Table 2 for the total population and for the low and high educated group. In the total study population, being obese or having a large waist circumference was associated with having an unhealthy diet (OR for obese compared to normal weight participants 1.31 [95% CI: 1.22; 1.40], OR for large waist compared to normal waist participants 1.32 [1.25; 1.40]). Significant interactions were found with educational level ($p<0.0001$): associations were stronger in the high educated group, although borderline significant associations in the same direction also existed in the low educated group. An inactive lifestyle, no daily breakfast and no dietary supplement use were all associated with an unhealthy diet. These associations were present both in low and high educated persons, although associations for dietary supplement use were slightly stronger in the highly educated group (OR 1.13 [1.01; 1.27] compared to 1.08 [0.96; 1.21] in the low educated group). Neighborhood social status score was not associated with an unhealthy diet, nor in the total group, neither in low or high educated persons. A significant interaction was found between urbanization level and educational level ($p=0.006$): among high educated participants those living in an extremely urbanized area less frequently reported an unhealthy diet than those living in a low urbanized area. However, this association did not reach statistical significance: OR 0.86 [0.71; 1.04]. No association was observed between urbanization level and an unhealthy diet in the low educated group. The association of smoking status with an unhealthy diet clearly differed by level of education (P interaction < 0.0001).

Table 1 Characteristics of the total EPIC-NL study population, and stratified by educational level

			Total population	Educational level	
			N=39,393	Low (N=6441)	High (N=8046)
Age, in years		Mean (SD)	49.2 (11.9)	56.0 (9.1)	46.5 (11.0)
Sex	Males	N (%)	10058 (26%)	1099 (17%)	2693 (33%)
	Females	N (%)	29335 (74%)	5342 (83%)	5353 (67%)
Prevalent disease	Yes	N (%)	3740 (9%)	966 (15%)	548 (7%)
	No	N (%)	35653 (91%)	5475 (85%)	7498 (93%)
BMI, kg/m^2		Mean (SD)	25.7 (4.0)	27.3 (4.4)	24.4 (3.3)
BMI categories	Normal	N (%)	18131 (46%)	1830 (28%)	4919 (61%)
	Overweight	N (%)	15671 (40%)	2989 (46%)	2607 (32%)
	Obese	N (%)	5591 (14%)	1622 (25%)	520 (6%)
Smoking status	Never	N (%)	15087 (38%)	2613 (41%)	3129 (39%)
	Former	N (%)	12318 (31%)	1714 (27%)	2750 (34%)
	Current	N (%)	11988 (30%)	2114 (33%)	2167 (27%)
Alcohol consumption	Grams ethanol per day	Median (IQR)	4.9 (15.0)	1.3 (8.2)	10.2 (18.2)
Physical activity level	Inactive	N (%)	3077 (8%)	878 (14%)	457 (6%)
	Moderately inactive	N (%)	9814 (25%)	1781 (28%)	1987 (25%)
	Moderately active	N (%)	10206 (26%)	1410 (22%)	2405 (30%)
	Active	N (%)	16296 (41%)	2372 (37%)	3197 (40%)
Dietary supplement use	No	N (%)	25804 (66%)	4524 (71%)	4789 (60%)
	Yes	N (%)	13338 (34%)	1855 (29%)	3220 (40%)
Frequency of breakfast consumption	Daily	N (%)	32388 (82%)	5469 (85%)	6485 (81%)
Unhealthy diet (mMDS<4)		N (%)	11770 (30%)	2507 (39%)	1620 (20%)
Working	Not working	N (%)	15357 (39%)	2153 (33%)	1623 (20%)
	Working	N (%)	24036 (61%)	4288 (67%)	6423 (80%)

Table 1 continued

			Total population	Educational level	
			N=39,393	Low (N=6441)	High (N=8046)
Urbanisation level [1]	Extremely	N (%)	10515 (28%)	1880 (31%)	2982 (39%)
	Strongly	N (%)	5948 (16%)	1032 (17%)	983 (13%)
	Moderately	N (%)	8056 (21%)	1252 (20%)	1446 (19%)
	Hardly	N (%)	6203 (16%)	894 (15%)	1110 (14%)
	Not	N (%)	6935 (18%)	1061 (17%)	1152 (15%)
Neighborhood Social status score [2]		Mean (SD)	-0.32 (0.85)	-0.47 (0.88)	-0.28 (0.89)

1Urbanization level: extremely urbanized (>2499 addresses/km2) , strongly urbanized (1500-2499 addresses/km2), moderately urbanized (1000-1499 addresses/km2) , hardly urbanized (500-999 addresses/km2), not urbanized (0-499 addresses/km2); 2 Social status score represents the social status of a neighborhood in 1998, based on: mean income, and the percentage of citizens that were unemployed, had a low income and had a low educational level, mean score in the Netherlands: -0.26. A higher score represents a higher social status of the neighborhood.

Among low educated persons, current smokers reported more frequently an unhealthy diet than never smokers (OR 1.42 [95% CI: 1.25; 1.61]), while among highly educated participants, smokers less frequently reported an unhealthy diet (OR 0.89 [0.78; 1.02]). Highly educated former smokers also less frequently reported an unhealthy diet than highly educated never smokers (OR 0.80 [0.70; 0.92]). Former smoking was not associated with an unhealthy diet in the group with a low education. Excluding participants with a BMI below 18.5 kg/m^2 did not alter the results, nor did excluding alcohol from the mMDS (Supplemental table 1).

Table 2 Odds ratios (95% CI) for having an unhealthy diet, in the total study population and stratified by educational level[1]

		Total population (N=39,393)		Educational level				
				Low (N=6379)		High (N=8009)		
Individual factors		OR (95% CI)	P for trend	OR (95% CI)	P for trend	OR (95% CI)	P for trend	Interaction with educational level[2]
BMI[3]	Normal weight	Reference	<0.0001	Reference	0.10	Reference	0.0003	<0.0001
	Overweight	1.09 (1.04; 1.15)		0.99 (0.88; 1.12)		1.09 (0.96; 1.24)		
	Obese	1.31 (1.22; 1.40)		1.14 (0.99; 1.31)		1.59 (1.29; 1.96)		
Waist circumference[3]	Normal	Reference	<0.0001	Reference	0.05	Reference	<0.0001	<0.0001
	Moderately increased	1.10 (1.04; 1.16)		1.06 (0.93; 1.22)		1.17 (1.02; 1.34)		
	Large	1.32 (1.25; 1.40)		1.14 (1.00; 1.30)		1.47 (1.25; 1.72)		
Smoking	Never	Reference	0.0029	Reference	<0.0001	Reference	0.0047	<0.0001
	Former	0.79 (0.75; 0.84)		0.98 (0.86; 1.12)		0.80 (0.70; 0.92)		
	Current	1.11 (1.05; 1.17)		1.42 (1.25; 1.61)		0.89 (0.78; 1.02)		
Physical activity	Active	Reference	<0.0001	Reference	0.0003	Reference	<0.0001	0.63
	Moderately active	1.04 (0.98; 1.10)		0.97 (0.84; 1.12)		1.12 (0.98; 1.29)		
	Moderately inactive	1.16 (1.10; 1.23)		1.21 (1.06; 1.39)		1.35 (1.17; 1.55)		
	Inactive	1.36 (1.25; 1.48)		1.26 (1.06; 1.49)		1.37 (1.08; 1.74)		

Table 2 continued

		Total population (N=39,393)		Educational level				
				Low (N=6379)		High (N=8009)		
Individual factors		OR (95% CI)	P for trend	OR (95% CI)	P for trend	OR (95% CI)	P for trend	Interaction with educational level[2]
Dietary supplement use	Yes	Reference		Reference		Reference		0.14
	No	1.22 (1.16; 1.28)		1.08 (0.96; 1.21)		1.13 (1.01; 1.27)		
Frequency of breakfast	Daily	Reference		Reference		Reference		0.75
	Less than daily	1.27 (1.20; 1.35)		1.29 (1.11; 1.49)		1.28 (1.12; 1.46)		
Working status	working	Reference		Reference		Reference		0.34
	not working	1.21 (1.15; 1.28)		1.05 (0.93; 1.19)		0.99 (0.85; 1.14)		
Neighborhood factors								
Urbanisation level[4]	Not	Reference	0.81	Reference	0.48	Reference	0.03	0.006
	Hardly	1.03 (0.88; 1.20)		1.03 (0.31; 3.45)		1.09 (0.86; 1.39)		
	Moderately	1.01 (0.89; 1.16)		0.95 (0.31; 2.91)		1.00 (0.80; 1.23)		
	Strongly	1.15 (1.00; 1.32)		1.01 (0.31; 3.27)		0.97 (0.77; 1.23)		
	Extremely	0.99 (0.88; 1.11)		1.06 (0.38; 2.98)		0.86 (0.71; 1.04)		
Neighborhood Social status score[4]	Quartile 4	Reference	0.028	Reference	0.21	Reference	0.47	0.11
	Quartile 3	1.03 (0.92; 1.15)		0.97 (0.35; 2.71)		1.04 (0.87; 1.24)		
	Quartile 2	1.06 (0.93; 1.21)		1.11 (0.38; 3.29)		1.00 (0.83; 1.22)		
	Quartile 1	1.11 (0.99; 1.24)		1.07 (0.39; 2.91)		0.94 (0.78; 1.13)		

[1] adjusted for age, gender, prevalent diseases, and the other individual factors. An unhealthy diet is defined as low mMDS (mMDS <4); [2] interaction term is calculated for a model including the low and high educational level only (N=14,388); [2] BMI and waist are not adjusted for each other; [4] results based on multi-level analysis.

DISCUSSION

We investigated whether associations between lifestyle factors and an unhealthy diet, defined as a low adherence to a Mediterranean-style diet, differ by educational level. Most associations were consistent across low and high educational level and comparable to those observed in the total study population. The association of smoking with an unhealthy diet clearly differed by educational level: opposite associations were found. Current smokers had a greater odds of an unhealthy diet compared with never smokers if they had a low educational level, but a lower odds of an unhealthy diet if they had a high educational level. Additionally, we found a lower odds of an unhealthy diet for former compared to never smokers among high educated, but not among low educated participants. Although also for BMI and waist circumference significant interactions with educational level were found, we did not observe different trends in low and high educated participants.

Other studies reported on a higher clustering of unhealthy behaviors in persons with a low socio-economic status (8, 23). In this study, we investigated if individual unhealthy behaviors were also related to having an unhealthy diet among high educated people. We found that most associations were consistent across participants with a low and high educational level. For smoking, opposite associations were found in participants with a low or high educational level. We do not know the reason for this result, but speculate that high educated people who smoke may be more aware of their unhealthy smoking behavior and compensate this (intentionally or unintentionally) by adapting other healthier behaviors, such as a healthy diet. A similar compensating effect may exist in high educated former smokers. High educated former smokers reported less often an unhealthy diet compared to never smokers, suggesting they adapted to an overall healthier lifestyle.

In other studies, a higher socio-economic status has been associated with a healthier dietary pattern (4, 24). Mullie *et al* investigated the relationship between socioeconomic status and three dietary patterns, using the Healthy Eating Index, the Mediterranean Diet Score and a pattern based on principal component analysis. Higher adherence to all three patterns was associated with a higher socioeconomic status (7). We defined a healthy dietary pattern by the modified Mediterranean Diet Score. Consistent with our results, a low educational level (5, 25), being physically inactive (5, 25) and smoking (5) have all been related to a lower adherence to a Mediterranean-style diet before. In contrast to our results, skipping breakfast was not significantly associated with adherence to a Mediterranean diet in a study on obesity-related eating behaviors (26). However, that study investigated the association with never having breakfast, while we

investigated less than daily breakfast frequency. Our finding that dietary supplement use is associated with having a healthier diet has also been observed in other studies (27, 28).

In our study, a score for neighborhood social status was not related to unhealthy diet, not in the total group, and not in low and high educated persons. The score was also not related to physical activity level and obesity. This can be due to the fact that there is no extreme socioeconomic distribution in the Netherlands (11), but perhaps also to a possible selective participation of the higher educated persons from neighborhoods with a low social status in the EPIC-NL study. We therefore concluded that we cannot use this score to differentiate between neighborhoods with healthy or unhealthy behaviors and thus did not use this score to investigate interactions of neighborhood social status with other lifestyle factors. Two recent studies in Australia and Finland did find an influence of neighborhood social status on the co-occurrence of unhealthy lifestyle factors, i.e. people in more affluent neighborhoods had less unhealthy lifestyle factors (9, 10). In both studies unhealthy lifestyles (i.e. smoking, alcohol use and physical activity level) were combined into an index, while we related neighborhood social status score to an unhealthy diet.

Strengths of our study are the use of a large population that included detailed information on diet, lifestyle factors and educational level. Several limitations need to be addressed as well. Dietary patterns were based on self-reported intake from a FFQ. Furthermore, detailed information on several individual and neighborhood factors was available in our study, but we lacked information on social-cultural and economic factors. These factors may also be associated with a healthy diet (29).

Conclusions

In conclusion, associations between lifestyle factors and unhealthy diet were consistent across educational levels. Differences were only found for current and former smokers: in low educated persons, current smoking was related to an unhealthy diet in comparison to never smoking, whereas in high educated persons current and former smoking was related to a healthier diet. These results can be used in the development of targeted health promotion strategies. If these strategies include smoking and diet, educational level of the target population should be taken into account.

REFERENCES

1. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369(10):954-64.

2. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2197-223.

3. Kant AK. Dietary patterns and health outcomes. J Am Diet Assoc. 2004;104(4):615-35.

4. Boylan S, Lallukka T, Lahelma E, Pikhart H, Malyutina S, Pajak A, et al. Socio-economic circumstances and food habits in Eastern, Central and Western European populations. Public Health Nutr. 2011; 14(4):678-87.

5. Hu EA, Toledo E, Diez-Espino J, Estruch R, Corella D, Salas-Salvado J, et al. Lifestyles and risk factors associated with adherence to the Mediterranean diet: a baseline assessment of the PREDIMED trial. PLoS One. 2013;8(4):e60166.

6. Patino-Alonso MC, Recio-Rodriguez JI, Belio JF, Colominas-Garrido R, Lema-Bartolome J, Arranz AG, et al. Factors associated with adherence to the Mediterranean diet in the adult population. J Acad Nutr Diet. 2014;114(4):583-9.

7. Mullie P, Clarys P, Hulens M, Vansant G. Dietary patterns and socioeconomic position. Eur J Clin Nutr. 2010; 64(3):231-8.

8. Pampel FC, Krueger PM, Denney JT. Socioeconomic Disparities in Health Behaviors. Annu Rev Sociol. 2010;36:349-70.

9. Feng X, Astell-Burt T. Neighborhood socioeconomic circumstances and the co-occurrence of unhealthy lifestyles: evidence from 206,457 Australians in the 45 and up study. PLoS One. 2013;8(8):e72643.

10. Halonen JI, Kivimaki M, Pentti J, Kawachi I, Virtanen M, Martikainen P, et al. Quantifying neighbourhood socioeconomic effects in clustering of behaviour-related risk factors: a multilevel analysis. PLoS One. 2012;7(3):e32937.

11. Giskes K, Turrell G, van Lenthe FJ, Brug J, Mackenbach JP. A multilevel study of socio-economic inequalities in food choice behaviour and dietary intake among the Dutch population: the GLOBE study. Public Health Nutr. 2006;9(1):75-83.

12. Sofi F, Abbate R, Gensini GF, Casini A. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. Am J Clin Nutr. 2010; 92(5):1189-96.

13. Fransen HP, Beulens JW, May AM, Struijk EA, Boer JMdW, G. A., Onland-Moret NC, et al. Dietary patterns in relation to quality-adjusted life years in the EPIC-NL cohort. Prev Med. 2015;Aug(77):119-24.

14. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8.

15. Ocke MC, Bueno-de-Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, et al. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol. 1997;26 Suppl 1:S37-S48.

16. Ocke MC, Bueno-de-Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol. 1997;26 Suppl 1:S49-58.

17. Voedingscentrum. Dutch food consumption table 1996.

18. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocke MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ. 2005;330(7498):991.

19. Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr. 2003;6(4):407-13.

20. WHO. Obesity: Preventing and managing the global epidemic. Report of a WHO Consultation (WHO Technical Report Series 894). Geneva: 2000.

21. Statistics Netherlands. Kerncijfers wijken en buurten 1995. Den Haag/Heerlen: CBS; [Accessed 20-11-2014]; Available from: http://statline.cbs.nl/.

22. Sociaal en Cultureel Planbureau. Statusscores. 2015; Available from: http://www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken/Statusscores.

23. Buck DF, F. Clustering of unhealthy behaviours over time. Implications for policy and practice. London: The King's Fund, 2012.

24. Darmon N, Drewnowski A. Does social class predict diet quality? Am J Clin Nutr. 2008;87(5):1107-17.

25. Leon-Munoz LM, Guallar-Castillon P, Graciani A, Lopez-Garcia E, Mesas AE, Aguilera MT, et al. Adherence to the Mediterranean diet pattern has declined in Spanish adults. J Nutr. 2012;142(10):1843-50.

26. Mesas AE, Guallar-Castillon P, Leon-Munoz LM, Graciani A, Lopez-Garcia E, Gutierrez-Fisac JL, et al. Obesity-related eating behaviors are associated with low physical activity and poor diet quality in Spain. J Nutr. 2012;142(7):1321-8.

27. Kirk SF, Cade JE, Barrett JH, Conner M. Diet and lifestyle characteristics associated with dietary supplement use in women. Public Health Nutr. 1999;2(1):69-73.

28. Foote JA, Murphy SP, Wilkens LR, Hankin JH, Henderson BE, Kolonel LN. Factors associated with dietary supplement use among healthy adults of five ethnicities: the Multiethnic Cohort Study. Am J Epidemiol. 2003;157(10):888-97.

29. Brug J. Determinants of healthy eating: motivation, abilities and environmental opportunities. Fam Pract. 2008;25 Suppl 1:i50-5.

Supplemental table 1 Sensitivity analysis: odds ratios (95% CI) for having an unhealthy diet , in the total study population and stratified by educational level[1]

		Total population (N=39,393)		Educational level				
				Low (N=6379)		High (N=8009)		Interaction with educational level[2]
		OR (95% CI)	P for trend	OR (95% CI)	P for trend	OR (95% CI)	P for trend	
Individual factors								
BMI [3]	Normal weight	Reference	<0.0001	Reference	0.20	Reference	0.005	0.05
	Overweight	1.07 (1.02; 1.12)		0.96 (0.85; 1.08)		1.10 (0.98; 1.22)		
	Obese	1.19 (1.11; 1.27)		1.11 (0.96; 1.27)		1.33 (1.09; 1.62)		
Waist circumference [3]	Normal	Reference	<0.0001	Reference	0.20	Reference	0.0002	0.22
	Moderately increased	1.09 (1.04; 1.15)		1.03 (0.90; 1.18)		1.18 (1.05; 1.33)		
	Large	1.22 (1.16; 1.28)		1.09 (0.96; 1.23)		1.29 (1.12; 1.49)		
Smoking	Never	Reference	<0.0001	Reference	<0.0001	Reference	0.66	<0.001
	Former	0.93 (0.88; 0.97)		1.10 (0.97; 1.24)		0.94 (0.84; 1.06)		
	Current	1.24 (1.18; 1.30)		1.59 (1.40; 1.80)		1.04 (0.92; 1.17)		
Physical activity	Active	Reference	<0.0001	Reference	0.005	Reference	<0.0001	0.16
	Moderately active	1.07 (1.01; 1.13)		0.94 (0.82; 1.08)		1.17 (1.04; 1.32)		
	Moderately inactive	1.19 (1.13; 1.26)		1.14 (1.00; 1.31)		1.36 (1.20; 1.54)		
	Inactive	1.34 (1.24; 1.45)		1.19 (1.01; 1.41)		1.38 (1.11; 1.71)		

Supplemental table 1 continued

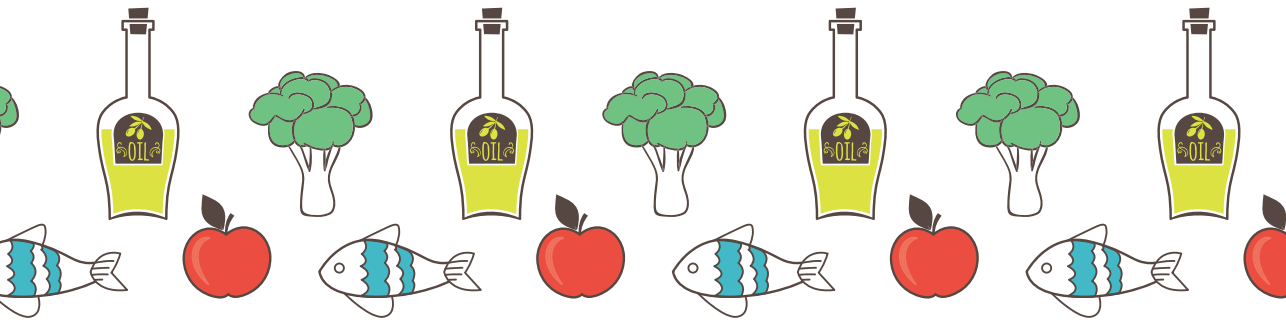
		Total population (N=39,393)		Educational level				
				Low (N=6379)		High (N=8009)		Interaction with educational level[2]
		OR (95% CI)	P for trend	OR (95% CI)	P for trend	OR (95% CI)	P for trend	
Individual factors								
Dietary supplement use	Yes	Reference		Reference		Reference		0.02
	No	1.22 (1.16; 1.27)		1.05 (0.93; 1.17)		1.18 (1.07; 1.31)		
Frequency of breakfast	Daily	Reference		Reference		Reference		0.57
	Less than daily	1.33 (1.26; 1.40)		1.30 (1.13; 1.51)		1.27 (1.12; 1.43)		
Working status	working	Reference		Reference		Reference		0.56
	not working	1.10 (1.05; 1.15)		1.00 (0.88; 1.12)		0.96 (0.84; 1.09)		
Neighborhood factors								
Urbanisation level [4]	Not	Reference	0.68	Reference	0.33	Reference	0.17	0.01
	Hardly	1.02 (0.89; 1.17)		1.11 (0.90; 1.36)		1.02 (0.32; 3.30)		
	Moderately	1.04 (0.92; 1.17)		1.09 (0.90; 1.31)		0.94 (0.31; 2.83)		
	Strongly	1.10 (0.98; 1.25)		1.06 (0.97; 1.29)		0.91 (0.27; 3.10)		
	Extremely	0.97 (0.87; 1.07)		1.12 (0.94; 1.33)		0.80 (0.29; 2.17)		
Neighborhood Social status score [4]	Quartile 4	Reference	0.32	Reference	0.34	Reference	0.06	0.51
	Quartile 3	0.98 (0.89; 1.09)		0.97 (0.82; 1.15)		0.94 (0.82; 1.09)		
	Quartile 2	1.03 (0.92; 1.16)		1.08 (0.91; 1.30)		0.94 0.81; 1.10)		
	Quartile 1	1.03 (0.93; 1.14)		1.05 (0.89; 1.23)		0.86 (0.74; 1.00)		

[1] adjusted for age, gender, prevalent diseases, and the other individual factors. An unhealthy diet is defined as low mMDS (mMDS excluding alcohol <4); [2] interaction term is calculated for a model including the low and high educational level only (N=14,388); [3] BMI and waist are not adjusted for each other; [4] results based on multi-level analysis.

Chapter 7

Exposure to famine at a young age and unhealthy lifestyle behavior later in life

Based on: Heidi P Fransen, Petra HM Peeters, Joline WJ Beulens, Jolanda MA Boer, G Ardine de Wit, N Charlotte Onland-Moret, Yvonne T van der Schouw, H Bas Bueno-de-Mesquita, Jeljer Hoekstra, Sjoerd G Elias, Anne M May.
Exposure to famine at a young age and unhealthy lifestyle behavior later in life.
Submitted for publication.



ABSTRACT

Introduction: A healthy diet is important for normal growth and development. Exposure to undernutrition during important developmental periods such as childhood and adolescence can have effects later in life. Inhabitants of the west of the Netherlands were exposed to severe undernutrition during the famine in the last winter of the second World War (1944-1945). We investigated if exposure of women to the Dutch famine during childhood and adolescence was associated with an unhealthy lifestyle later in life.

Methods: We studied 7,525 women from the Prospect-EPIC cohort, recruited in 1993-97 and aged 0-18 years during the Dutch famine. An individual famine score was calculated based on self-reported information about experience of hunger and weight loss. We investigated the association between famine exposure in early life and four lifestyle factors in adulthood: smoking, alcohol consumption, physical activity level and a Mediterranean-style diet.

Results: Of the 7,525 included women, 46% were unexposed, 38% moderately exposed and 16% severely exposed to the Dutch famine. Moderately and severely exposed women were more often former or current smokers compared to women that did not suffer from the famine: adjusted prevalence ratio 1.10 (95% CI: 1.05; 1.14) and 1.18 (1.12; 1.25), respectively. They also smoked more pack years than unexposed women. Severely exposed women were more often physically inactive than unexposed women, adjusted prevalence ratio 1.32 (1.06; 1.64). Results did not differ between exposure age categories (0-9 and 10-17 years). We found no associations of famine exposure with alcohol consumption and no dose-dependent relations with diet.

Conclusion: Exposure to famine early in female life may be associated with higher prevalence of smoking and physical inactivity later in life, but not with unhealthy diet and alcohol consumption.

INTRODUCTION

A healthy diet is important for normal growth and development, especially during important developmental periods such as childhood and adolescence (1). The developmental origins of health and disease hypothesis posits that undernutrition during fetal and infant life results in early adaptations of the body, which may lead to chronic disease later in life (2). This hypothesis is supported by results from Dutch famine studies (3-6).

The Dutch famine took place in the winter of 1944-1945. Inhabitants of the Western part of the Netherlands were exposed to severe undernutrition in the last 6 months of the Second World War. This historical event created a unique opportunity to gain insight into the long-term effects of a relatively short period of transient undernutrition. Because of the short exposure period, it is possible to pinpoint effects to specific growth periods in human life.
Increased risks of overweight, diabetes, coronary heart disease, COPD and asthma have been reported in individuals who were exposed to the Dutch famine (3-6). Furthermore, famine exposure was associated with an increased risk of breast cancer in one study (7), while others found no clear effects (8). No associations were found with non-breast cancer risk (9). The associations between famine exposure early in life and various biological outcomes may be due to biological effects, i.e. epigenetic (10) or hormonal changes (11), or to behavioral reactions following the exposure. The association between undernutrition early in life and different health behaviors later in life has not been investigated in depth before. To the best of our knowledge only one working paper describes the association between undernutrition and dietary intake. Kesternich *et al.* suggested that early-life shocks affect nutritional behavior later in life (12). Exposure to hunger during childhood was related to an increased fraction of income that was spent on food later in life. However, true food intake was not measured and it was therefore not known if they consumed healthy or unhealthy products. No studies on other lifestyle factors are available. Unhealthy behaviors, such as smoking, drinking, being physically inactive, and eating an unhealthy diet, are important risk factors for many non-communicable diseases (13, 14) and may act as an intermediate factor between famine exposure and chronic disease occurrence later in life. In the present study we therefore investigate if exposure to the Dutch famine during childhood and adolescence is associated with an unhealthy lifestyle later in life. We focus on the lifestyle factors smoking, alcohol consumption, physical activity level and usual diet.

METHODS

The Dutch famine

During the Second World War, from October 1944 till April 1945, inhabitants of the occupied Western part of the Netherlands were exposed to famine. Their daily food rations dropped to less than 25% of the pre-famine rations and varied between 400-800 kcal/day (15). After approximately 6 months of hunger the famine ended abruptly by liberation of the Netherlands in May 1945, and food became available again through supplies of the allied forces. This short period of extreme hunger allows the study of long-term effects of famine exposure.

The Prospect-EPIC cohort

We investigated the association between famine exposure and an unhealthy lifestyle in the Prospect-EPIC cohort. This is one of two Dutch cohorts of the European Prospective Investigation into Cancer and Nutrition (16, 17). Between 1993 and 1997 17,357 women were recruited in the Prospect-EPIC cohort. They all participated in the nationwide breast cancer screening program and were living in the city of Utrecht or surroundings. At recruitment, the women completed a general questionnaire (containing among others three questions about exposure to the 1944-1945 famine) and a validated food frequency questionnaire (FFQ) (18, 19), and underwent a physical examination. All participants provided written informed consent before study inclusion. The Prospect-EPIC study complies with the Declaration of Helsinki and was approved by the Institutional Review Board of the University Medical Center Utrecht.

Exclusion criteria

We excluded participants who answered 'not applicable' or 'I don't know' to one or more of the three famine exposure questions (n = 4975). Furthermore, we excluded women who were born after the Dutch famine (n=2559) or who were >18 years during the famine (N=481), or who lived outside the Netherlands during the famine (n=1732), or who had no dietary information available (n=85). Our final study population consisted of 7,525 women.

Individual famine score

Participants were asked about their experience of hunger and weight loss during the famine (3). The questions each contained the answer categories 'hardly', 'little', and 'very much'. These categories were combined into a three-point famine exposure score,

as previously reported: 1) severely exposed: women who reported being 'very much' exposed to both hunger and weight loss; 2) unexposed: women who reported 'hardly' being exposed to both hunger and weight loss; and 3) moderately exposed: all others (3).

Exposure age categories

We divided women into two age categories, using age at start of the famine (October 1st, 1944), because we wanted to investigate the effect of famine exposure during different growth periods. These categories were made according to the human life cycle as defined by Bogin (20) and have been used in the Prospect-EPIC cohort before (4) : 0-9 years (childhood, n=4385) , and 10-17 years (adolescence, n=3140).

Unhealthy lifestyle factors

Smoking

Information on smoking status and smoking intensity was available from the general questionnaire at recruitment (1993-7). Smoking status was defined as current, former or never smoker (categorical). Pack years of smoking (continuous) were calculated as packs (25 cigarettes) smoked per day multiplied by years of smoking. Pack years were available and analyzed for current and former smokers.

Alcohol consumption

Information on alcohol consumption from the baseline questionnaire (being a never, former or current drinker) was combined with alcohol intake from the food frequency questionnaire (in grams ethanol per day) and categorized into never drinkers (abstainers), light current drinkers (>0-5 g/day), moderate current drinkers (5-15 g/day) and heavy current drinkers (≥15 g/day). Furthermore, the amount of alcohol intake was analyzed among women who drank ≥1 g/day. For women who drank less, their intake may come from other products than alcoholic drinks, i.e. chocolate candy or sauces.

Physical activity level

Physical activity level was assessed in the general questionnaire and categorized according to the validated Cambridge Physical Activity Index into inactive, moderately inactive, moderately active or active (21).

Diet

The modified Mediterranean Diet Score (mMDS) was used as a measure of a healthy diet (22). Compared with the original Mediterranean Diet Score fish and poly-unsaturated fatty acids were additionally included in this score (23). A high score is associated with lower risk of chronic diseases (24) and in the total EPIC-NL cohort with a longer healthy life expectancy (25). Information of the food frequency questionnaire was used to score intake of eight components of the mMDS: vegetables; legumes; fruit, nuts and seeds; cereals; fish; the ratio of unsaturated to saturated fatty acids; meat; and dairy products. For the first 6 components intake equal to or above the study population median was assigned a value of 1, and intake below the median a value of 0. For meat and dairy products intake equal to or below the median was assigned a value of 1. Points were summed into the modified Mediterranean Diet Score, ranging from zero to eight points We did not include alcohol consumption in the score, as alcohol consumption was investigated as a separate lifestyle factor. A low self-reported modified Mediterranean Diet Score, i.e. a score below 4, was defined as an unhealthy diet. Furthermore, the score was analyzed continuously.

Covariates

We used age at start of the famine (1st October 1944) and educational level, which is considered to be a proxy for socioeconomic status, as covariates in our analyses. We categorized levels of education into very low (only primary school), low (lower vocational education), middle (secondary school or intermediate vocational training) and high education (higher vocational training or university). Next, body mass index (BMI) and energy intake (kcal/day) were included as covariates. BMI (kg/m^2) was calculated from measured weight and height and used as a continuous variable. Energy intake was calculated in kcal/day using food frequency questionnaire data; and used as a continuous variable. For smoking as a covariate, smoking status and intensity were combined and categorized into 8 categories, i.e. current smoker (<15 cigarettes/day, 15-25 cigarettes/day, >25 cigarettes a day, pipe of cigar smoker), former smoker (quit <10 year ago, quit 10-20 year ago, quit >20 year ago) and never smoker.

Statistical analysis

Missing physical activity data were imputed using single imputation as previously described (26). Missing data on BMI and educational level were imputed, using single imputation regression modelling (SPSS-MVA). Characteristics of the study population are presented according to level of famine exposure as mean and standard deviation or as a percentage. Associations between famine exposure and lifestyle were determined for

the total study population and by age category. For categorical variables, we used a Poisson regression model, because an odds ratio will overestimate the effect size because of the high prevalences (27). Prevalence ratios were estimated for smoking (ever smokers [current or former smokers], vs. never smokers), drinking (heavy alcohol consumption, ≥15 g/day, vs. drinking less than 15 g/day), having an unhealthy diet (mMDS below 4 vs. mMDS of 4 or higher) and physical inactivity (inactive vs the rest, the latter includes participants that were active, moderately active and moderately inactive). The associations between famine exposure and continuous variables, i.e. pack years of smoking (only for current and former smokers), alcohol intake (in grams ethanol/day; only for current drinkers that drink >1 g/day), and modified Mediterranean Diet Score were estimated by linear regression.

Crude and multivariable models are presented. Multivariable models are adjusted for 1) age at start of the famine (October 1st, 1944; continuous) and educational level (categorical); 2) all variables in model 1 and BMI, energy intake, physical activity level, alcohol/smoking status and intensity, and mMDS. Covariates were excluded in the analyses where they are the outcome. A P value for linear trend was computed by including the categorical famine exposure score as a continuous variable in the model. We studied whether results were comparable for the two exposure age categories (0-9 and 10-17 years) by including interaction terms in the model. All statistical analyses were conducted using SAS 9.2 (SAS Institute, Cary, US). P-values <0.05 were considered to be statistically significant.

RESULTS

Characteristics of the study population according to self-reported level of famine exposure are presented in table 1. Of the 7,525 included women, 46% were unexposed, 38% moderately exposed and 16% severely exposed to famine. Participants who were severely exposed to famine were somewhat older at start of the famine and more often lower educated.

Associations between famine exposure and unhealthy lifestyle factors are presented for the total study population in table 2 and 3. Results stratified by age category (0-9 years or 10-17 years at start of the famine) are presented in the supplemental tables. Famine exposure was dose-dependently associated with prevalence of ever smoking; adjusted prevalence ratios were 1.10 (95% CI: 1.05; 1.14) and 1.18 (1.12; 1.25) for moderately exposed and severely exposed women, respectively (P for trend <0.0001) (table 2). This was observed in both age categories (supplemental table 1).

Table 1 Characteristics of the study population at recruitment, according to level of famine exposure, n=7,525.

			Level of famine exposure		
			Unexposed	Moderately exposed	Severely exposed
Participants	N (%)		3450 (46%)	2838 (38%)	1237 (16%)
Age at start of famine (Oct 1st, 1944), in years	Mean (SD)		8.0 (5.3)	8.8 (5.4)	9.1 (5.1)
Aged 0-9 years during famine (childhood)	N (%)		2122 (62%)	1601 (56%)	662 (54%)
Aged 10-18 years during famine (adolescent)	N (%)		1328 (38%)	1237 (44%)	575 (46%)
Age at recruitment (1993-1997), in years	Mean (SD)		58.8 (5.4)	59.5 (5.5)	59.7 (5.2)
BMI, kg/m^2	Mean (SD)		26.0 (3.9)	26.2 (4.0)	26.2 (4.2)
Waist, cm	Mean (SD)		83.5 (9.8)	84.4 (9.9)	84.7 (10.4)
Level of education	N (%)	Very low	820 (24%)	627 (22%)	320 (26%)
		Low	1691 (49%)	1355 (48%)	617 (50%)
		Middle	461 (13%)	402 (14%)	152 (12%)
		High	478 (14%)	454 (16%)	148 (12%)
Smoking status	N (%)	Never	1667 (48%)	1246 (44%)	503 (41%)
		Former	1121 (32%)	997 (35%)	457 (37%)
		Current	662 (19%)	595 (21%)	277 (22%)
Physical activity level	N (%)	Inactive	206 (6%)	218 (8%)	112 (9%)
		Moderately inactive	951 (28%)	742 (26%)	288 (23%)
		Moderately active	875 (25%)	728 (26%)	339 (27%)
		Active	1418 (41%)	1150 (41%)	498 (40%)
Alcohol consumption	N (%)	Never	15 (0%)	17 (0%)	12 (1%)
		Light (0-5 g/day)	1917 (56%)	1562 (55%)	749 (61%)
		Moderate (5-15 g/day)	790 (23%)	682 (24%)	228 (18%)
		Heavy (≥ 15 g/day)	728 (21%)	577 (20%)	248 (20%)
Energy intake in kcal/day	Mean (SD)		1800 (420)	1790 (411)	1756 (428)
mMDS, excluding alcohol	Mean (SD)		4.0 (1.5)	4.1 (1.5)	4.0 (1.5)
Unhealthy diet (mMDS<4)	N (%)		1300 (38%)	979 (35%)	474 (38%)

Famine exposure was also dose-dependently associated with pack years of smoking; moderately exposed women reported smoking 0.98 (0.10; 1.87) more pack years than unexposed women, while severely exposed women reported 2.53 (1.39; 3.66) more pack years (P for trend <0.0001) (table 3). Results were similar across age categories (supplemental table 2), with no significant interaction of famine exposure with age category (P=0.51).

No association was found between famine exposure and heavy drinking: adjusted prevalence ratios for moderately and severely exposed women compared to unexposed women were 0.94 (0.85; 1.03) and 0.95 (0.84; 1.07), respectively (table 2). Although a significant interaction with age was found (P=0.04), in both age categories prevalence ratios were not statistically significant. Famine exposure was not associated with alcohol intake in grams per day, neither in the total population (table 3), nor in the two age categories (supplemental table 4).
Moderately exposed women less often reported having an unhealthy diet than unexposed women: adjusted prevalence ratio 0.92 (0.86; 0.98) (table 2). No differences were found between severely exposed and unexposed women. No significant interaction with age was observed (P=0.51) (supplemental table 5). We also investigated the mMDS continuously. In the total population, moderately exposed women had a 0.08 point (95% CI: 0.00; 0.16) higher mMDS, compared to unexposed women (table 3). No differences were found between severely exposed and unexposed women, and no interaction with age was found (P=0.77) (supplemental table 6).

Famine exposure was associated with physical inactivity. Both moderately exposed and severely exposed women were more often physically inactive than unexposed women, adjusted prevalence ratio 1.18 (0.99; 1.42) and 1.32 (1.06; 1.64), respectively (P for trend=0.08) (table 2). The dose-dependent relation was more pronounced in the older age category (P for trend=0.001) (supplemental table 7).

Table 2 Categorical analysis: prevalence ratios and 95% CI for smoking, drinking, an unhealthy diet, and physical inactivity, according to level of famine exposure.

Famine exposure level	Crude model	P for trend	Multivariable model 1[e]	P for trend	Multivariable model 2[e]	P for trend	Interaction with age
Smoking[a]							
Unexposed	Reference	<0.0001	Reference	<0.0001	Reference	<0.0001	0.23
Moderately	1.09 (1.04; 1.14)		1.09 (1.04; 1.14)		1.10 (1.05; 1.14)		
Severely	1.15 (1.09; 1.21)		1.17 (1.11; 1.24)		1.18 (1.12; 1.25)		
Drinking[b]							
Unexposed	Reference	0.37	Reference	0.88	Reference	0.24	0.04
Moderately	0.96 (0.87; 1.06)		0.97 (0.88; 1.06)		0.94 (0.85; 1.03)		
Severely	0.95 (0.84; 1.08)		1.01 (0.89; 1.14)		0.95 (0.84; 1.07)		
Unhealthy diet[c]							
Unexposed	Reference	0.60	Reference	0.55	Reference	0.26	0.51
Moderately	0.92 (0.86; 0.98)		0.92 (0.87; 0.99)		0.92 (0.86; 0.98)		
Severely	1.02 (0.94; 1.10)		1.01 (0.93; 1.09)		0.98 (0.91; 1.07)		
Physical inactivity[d]							
Unexposed	Reference	<0.0001	Reference	0.0008	Reference	0.081	0.32
Moderately	1.29 (1.07; 1.55)		1.23 (1.03; 1.48)		1.18 (0.99; 1.42)		
Severely	1.52 (1.22; 1.89)		1.42 (1.15; 1.77)		1.32 (1.06; 1.64)		

[a] being a former or current smoker; [b] heavy drinking, ≥15 g/day; [c] unhealthy diet is defined as mMDS<4 (excluding alcohol); [d] being physically inactive;
[e] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level;
multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, smoking status and intensity, alcohol consumption, and mMDS (covariates are excluded if they are the outcome).

Table 3 Continuous analysis of the association between famine exposure and pack years of smoking, alcohol consumption, and diet (regression coefficients and 95% CI)

Famine exposure level	N		Crude model	P for trend	Multivariable model 1[d]	P for trend	Multivariable model 2[d]	P for trend	Interaction with age
Pack years of smoking[a]		Packyears, mean (SD)							
Unexposed	1684	14.2 (12.9)	Reference	<0.0001	Reference	<0.0001	Reference	<0.0001	0.51
Moderately	1514	15.2 (13.6)	1.01 (0.08; 1.94)		0.95 (0.03; 1.87)		0.98 (0.10; 1.87)		
Severely	696	17.3 (14.1)	3.10 (1.92; 4.29)		2.58 (1.41; 3.75)		2.53 (1.39; 3.66)		
Alcohol intake[b]		Ethanol, g/day mean (SD)							
Unexposed	2360	12.8 (13.0)	Reference	0.78	Reference	0.87	Reference	0.36	0.50
Moderately	1949	12.7 (12.7)	-0.15 (-0.93; 0.64)		-0.16 (-0.93; 0.62)		-0.41 (-1.15; 0.33)		
Severely	783	12.7 (14.1)	-0.09 (-1.15; 0.96)		0.20 (-0.85; 1.25)		-0.32 (-1.32; 0.67)		
Diet[c]		mMDS, mean (SD)							
Unexposed	3450	4.0 (1.5)	Reference	0.33	Reference	0.31	Reference	0.10	0.77
Moderately	2838	4.1 (1.5)	0.10 (0.03; 0.17)		0.08 (0.01; 0.16)		0.09 (0.02; 0.17)		
Severely	1237	4.0 (1.5)	0.00 (-0.09; 0.10)		0.02 (-0.08; 0.11)		0.05 (-0.05; 0.14)		

[a] includes former and current smokers only; [b] only current drinkers that drink >1 g/day; [c] modified Mediterranean Diet Score excluding alcohol;
[d] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level;
multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, smoking status and intensity, alcohol consumption,and mMDS (covariates are excluded if they are the outcome).

DISCUSSION

In our study, women who reported severe exposure to famine during their youth were more often smokers and smoked more later in life compared to women who were not exposed. Exposed women were also more often physically inactive. Associations were dose-dependent: stronger exposure to famine was associated with higher prevalence of smoking and physical inactivity. We found no associations of famine exposure with alcohol consumption and no dose-dependent relations with diet.

These results are in accordance with our hypothesis that famine exposure during important developmental periods, such as childhood and adolescence may relate to an unhealthier lifestyle later in life. However, famine exposure was not associated with alcohol consumption later in life and no clear relations with the modified Mediterranean diet Score were found.

Relations of exposure to the Dutch famine with occurrence of chronic diseases later in life have been reported previously. Famine exposure was associated with higher rates of overweight, diabetes, coronary heart disease, COPD and asthma (3-6). These associations were only partly corrected for unhealthy behaviors. Unhealthy behaviors are important risk factors for these diseases (13, 14) and may act alone or in combination as intermediate factors between famine exposure and chronic disease occurrence later in life. Little information on the association between famine exposure and lifestyle later in life is available. Most studies focused on cognition, which is often related with lifestyle, in children following famine exposure (1, 28), and on prenatal instead of postnatal exposure.

Lussana *et al.* (29) investigated whether prenatal undernutrition modified lifestyle choices. In line with our results, exposed persons were less physically active compared to unexposed persons. They were also twice as likely to consume a high-fat diet. We used a healthy diet score, the modified Mediterranean Diet Score, to investigate the association with diet. We did not find an association between severe famine exposure and eating an unhealthy diet. However, moderately exposed women reported a higher modified Mediterranean Diet Score, indicating that these women ate a healthier diet than the unexposed women. We have no clear explanation for these results, which were especially present in the younger age category (0-9 years old during the famine). It has to be noted, however, that the moderately exposed group is a very diverse group. This group also contained women who were little exposed to hunger or weight loss, or very much exposed to either weight loss or hunger.

We observed a higher prevalence of smoking and physical inactivity in participants that were severely exposed to the Dutch famine. As we are the first to study this association, there are no other studies to compare our results with. We can only speculate about the biological pathways along which famine exposure early in life may be associated with lifestyle later in life. Severe undernutrition during important developmental periods might impair brain development, as rapid brain development takes place during pregnancy and in the first years of life (1). However, environment (brain development is affected by experience), timing of the exposure, the degree of exposure and the possibility of recovery might also influence long-term effects of famine exposure on brain function (1). Vucetic *et al.* studied the effect of early life protein restriction (pre- and postnatal) in mice and found behavioral abnormalities that were dopamine-related (30). Dopamine plays a role in behavioral responses, and a dysfunction of the dopamine system is associated with neurobehavioral disorders, like addiction. This may explain the results found for smoking, but not the fact that heavy alcohol consumption was not related to famine exposure. Apart from a direct effect of the famine, famine-related stress or war-related stress may play a role in the adoption of unhealthy behaviors (3, 12). Stressful life events have been associated with higher risk of smoking and drinking in adolescents (31).

The Prospect-EPIC cohort provided us with the unique opportunity to study the long-term relations of famine exposure with lifestyle behavior later in life. Strengths of our study are the large study population and the documented famine period. The exposure had a sudden onset and ending, and took place in a previously well-nourished population. In addition, we were able to calculate individual famine scores from our questionnaire. Usually, place of residence is used as a marker of famine exposure in other studies, which is less accurate. It has to be noted, however, that our individual famine scores are based on self-reported data on the experience of weight loss and hunger, and especially women in the youngest age category may depend on information from their family for the recollection of their exposure. However, in our study, participants who were severely exposed to the famine were older than participants that were moderately exposed or unexposed. This complies with the historical fact that young children were relatively protected from hunger during the war and supports the quality of our data (15). Our study population included women who participated in the Dutch national breast cancer screening program (participation rate around 78%) (32), resulting in possible selective participation of women with an overall healthier lifestyle (33). Results are also conditional on survival until recruitment into the Prospect-EPIC study (1993-7). It is possible that women with the unhealthiest lifestyles had already died

before the study started. This may have biased our results. Furthermore, as our study cohort only included women, we do not know if results will be similar for men.

In conclusion, this is the first study that investigated the association between a short period of extreme hunger in early life and the presence of unhealthy lifestyle factors later in life. In women, exposure to famine was associated with a higher prevalence of smoking and physical inactivity, while no clear relations were found with diet and alcohol consumption. Our results imply that, next to having direct biological effects that increase chronic disease risk, famine exposure might indirectly relate to chronic disease risk through unhealthy lifestyle factors.

REFERENCES

1. Prado EL, Dewey KG. Nutrition and brain development in early life. Nutr Rev. 2014;72(4):267-84.

2. Barker DJ. The fetal and infant origins of adult disease. BMJ. 1990; 301(6761):1111.

3. van Abeelen AF, Elias SG, Bossuyt PM, Grobbee DE, van der Schouw YT, Roseboom TJ, et al. Famine exposure in the young and the risk of type 2 diabetes in adulthood. Diabetes. 2012;61(9):2255-60.

4. van Abeelen AF, Elias SG, Bossuyt PM, Grobbee DE, van der Schouw YT, Roseboom TJ, et al. Cardiovascular consequences of famine in the young. Eur Heart J. 2012;33(4):538-45.

5. van Abeelen AF, Elias SG, Roseboom TJ, Bossuyt PM, van der Schouw YT, Grobbee DE, et al. Postnatal acute famine and risk of overweight: the dutch hungerwinter study. Int J Pediatr. 2012;2012:936509.

6. van Abeelen AF, Elias SG, de Jong PA, Grobbee DE, Bossuyt PM, van der Schouw YT, et al. Famine in the young and risk of later hospitalization for COPD and asthma. PLoS One. 2013;8(12):e82636.

7. Elias SG, Peeters PH, Grobbee DE, van Noord PA. Breast cancer risk after caloric restriction during the 1944-1945 Dutch famine. J Natl Cancer Inst. 2004;96(7):539-46.

8. Dirx MJ, van den Brandt PA, Goldbohm RA, Lumey LH. Diet in adolescence and the risk of breast cancer: results of the Netherlands Cohort Study. Cancer Causes Control. 1999;10(3):189-99.

9. Elias SG, Peeters PH, Grobbee DE, van Noord PA. The 1944-1945 Dutch famine and subsequent overall cancer incidence. Cancer Epidemiol Biomarkers Prev. 2005;14(8):1981-5.

10. Hughes LA, van den Brandt PA, de Bruine AP, Wouters KA, Hulsmans S, Spiertz A, et al. Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms. PLoS One. 2009;4(11):e7951.

11. Elias SG, Keinan-Boker L, Peeters PH, Van Gils CH, Kaaks R, Grobbee DE, et al. Long term consequences of the 1944-1945 Dutch famine on the insulin-like growth factor axis. Int J Cancer. 2004;108(4):628-30.

12. Kesternich IS, B.; Smith, J.P.; Winter, J.K. Working paper: Individual Behavior as a Pathway Between Early-Life Shocks and Adult Health: Evidence from Hunger Episodes in Post-War Germany. RAND Labor and Population, 2013 september 2013. Contract No.: WR-1015.

13. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369(10):954-64.

14. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2197-223.

15. Burger GCES, H.R.; Drummond, J.C. Malnutrition and starvation in Western Netherlands, September 1944 to July 1945. Part I and II. The Hague: General State Printing Office; 1948.

16. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14.

17. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8.

18. Ocke MC, Bueno-de-Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, et al. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol. 1997;26 Suppl 1:S37-S48.

19. Ocke MC, Bueno-de-Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol. 1997;26 Suppl 1:S49-S58.

20. Bogin B. Patterns of human growth . Cambridge: Cambridge University Press; 1999.

21. Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr. 2003;6(4):407-13.

22. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocke MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ. 2005;330(7498):991.

23. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, et al. Diet and overall survival in elderly people. BMJ. 1995; 311(7018):1457-60.

24. Sofi F, Abbate R, Gensini GF, Casini A. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. Am J Clin Nutr. 2010; 92(5):1189-96.

25. Fransen HP, Beulens JW, May AM, Struijk EA, Boer JMdW, G. A., Onland-Moret NC, et al. Dietary patterns in relation to quality-adjusted life years in the EPIC-NL cohort. Prev Med. 2015;Aug(77):119-24.

26. Joosten MM, Grobbee DE, van der AD, Verschuren WM, Hendriks HF, Beulens JW. Combined effect of alcohol consumption and lifestyle behaviors on risk of type 2 diabetes. Am J Clin Nutr. 2010;91(6):1777-83.

27. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159(7):702-6.

28. Ampaabeng SK, Tan CM. The long-term cognitive consequences of early childhood malnutrition: the case of famine in Ghana. J Health Econ. 2013;32(6):1013-27.

29. Lussana F, Painter RC, Ocke MC, Buller HR, Bossuyt PM, Roseboom TJ. Prenatal exposure to the Dutch famine is associated with a preference for fatty foods and a more atherogenic lipid profile. Am J Clin Nutr. 2008;88(6):1648-52.

30. Vucetic Z, Totoki K, Schoch H, Whitaker KW, Hill-Smith T, Lucki I, et al. Early life protein restriction alters dopamine circuitry. Neuroscience. 2010;168(2):359-70.

31. Simantov E, Schoen C, Klein JD. Health-compromising behaviors: why do adolescents smoke or drink?: identifying underlying risk and protective factors. Arch Pediatr Adolesc Med. 2000;154(10):1025-33.

32. Fracheboud J, de Koning HJ, Beemsterboer PM, Boer R, Hendriks JH, Verbeek AL, et al. Nation-wide breast cancer screening in The Netherlands: results of initial and subsequent screening 1990-1995. National Evaluation Team for Breast Cancer Screening. Int J Cancer. 1998;75(5):694-8.

33. Struijk EA, May AM, Beulens JW, van Gils CH, Monninkhof EM, van der Schouw YT, et al. Mortality and cancer incidence in the EPIC-NL cohort: impact of the healthy volunteer effect. Eur J Public Health. 2015;25(1):144-9.

Supplemental table 1 Prevalence ratios and 95% CI for smoking status[a], according to level of famine exposure, stratified by age category.

Age category and famine exposure level	Crude model	P for trend	Multivariable model 1[b]	P for trend	Multivariable model 2[b]	P for trend
All ages						
Unexposed	Reference	<0.0001	Reference	<0.0001	Reference	<0.0001
Moderately	1.09 (1.04; 1.14)		1.09 (1.04; 1.14)		1.10 (1.05; 1.14)	
Severely	1.15 (1.09; 1.21)		1.17 (1.11; 1.24)		1.18 (1.12; 1.25)	
0-9 years						
Unexposed	Reference	0.0040	Reference	0.0013	Reference	0.0002
Moderately	1.06 (1.00; 1.12)		1.07 (1.01; 1.13)		1.07 (1.01; 1.13)	
Severely	1.10 (1.03; 1.18)		1.12 (1.04; 1.20)		1.13 (1.06; 1.21)	
10-17 years						
Unexposed	Reference	<0.0001	Reference	<0.0001	Reference	<0.0001
Moderately	1.15 (1.06; 1.25)		1.14 (1.06; 1.24)		1.14 (1.06; 1.24)	
Severely	1.26 (1.15; 1.38)		1.24 (1.13; 1.36)		1.24 (1.13; 1.35)	

[a] being a former or current smoker;
[b] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level;
multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, alcohol consumption, and mMDS.

Supplemental table 2 Association between famine exposure and pack years of smoking, stratified by age category, regression coefficients and 95% CI, n=3,894[a]

Age category and famine exposure level	N	Packyears, mean (SD)	Crude model	P for trend	Multivariable model 1[b]	P for trend	Multivariable model 2[b]	P for trend
All ages								
Unexposed	1684	14.2 (12.9)	Reference	<0.0001	Reference	<0.0001	Reference	<0.0001
Moderately	1514	15.2 (13.6)	1.01 (0.08; 1.94)		0.95 (0.03; 1.87)		0.98 (0.10; 1.87)	
Severely	696	17.3 (14.1)	3.10 (1.92; 4.29)		2.58 (1.41; 3.75)		2.53 (1.39; 3.66)	
0-9 years								
Unexposed	1115	13.7 (12.5)	Reference	<0.0001	Reference	0.0007	Reference	0.0004
Moderately	903	14.4 (12.4)	0.63 (-0.48; 1.73)		0.64 (-0.44; 1.73)		0.72 (-0.33; 1.76)	
Severely	386	16.9 (13.1)	3.16 (1.70; 4.61)		2.65 (1.21; 4.09)		2.65 (1.26; 4.04)	
10-17 years								
Unexposed	569	15.2 (13.7)	Reference	<0.0001	Reference	0.0076	Reference	0.0118
Moderately	611	16.5 (15.1)	1.35 (-0.32; 3.02)		1.47 (-0.18; 3.12)		1.49 (-0.11; 3.10)	
Severely	310	17.9 (15.3)	2.69 (0.67; 4.71)		2.62 (0.63; 4.60)		2.37 (0.42; 4.31)	

[a] includes current and former smokers only;
[b] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level;
multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, alcohol consumption, and mMDS.

Supplemental table 3 Prevalence ratios and 95% CI for being a heavy drinker[a], according to level of famine exposure, stratified by age category.

Age category and famine exposure level	Crude model	P for trend	Multivariable model 1[b]	P for trend	Multivariable model 2[b]	P for trend
All ages						
Unexposed	Reference	0.37	Reference	0.88	Reference	0.24
Moderately	0.96 (0.87; 1.06)		0.97 (0.88; 1.06)		0.94 (0.85; 1.03)	
Severely	0.95 (0.84; 1.08)		1.01 (0.89; 1.14)		0.95 (0.84; 1.07)	
0-9 years						
Unexposed	Reference	0.81	Reference	0.82	Reference	0.57
Moderately	0.98 (0.87; 1.10)		0.97 (0.87; 1.09)		0.94 (0.84; 1.05)	
Severely	0.99 (0.85; 1.15)		1.04 (0.89; 1.21)		0.98 (0.85; 1.14)	
10-17 years						
Unexposed	Reference	0.84	Reference	0.56	Reference	0.17
Moderately	1.00 (0.83; 1.19)		0.95 (0.79; 1.13)		0.91 (0.76; 1.08)	
Severely	0.98 (0.77; 1.23)		0.94 (0.75; 1.19)		0.87 (0.69; 1.08)	

[a] heavy drinking, ≥15 g/day

[b] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level; multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, smoking status and intensity, and mMDS.

Supplemental table 4 Association between famine exposure and alcohol intake in grams ethanol per day[a], stratified by age category, regression coefficients and 95% CI, n=5,092.

Age category and famine exposure level	N	Mean (SD)	Crude model	P for trend	Multivariable model 1[b]	P for trend	Multivariable model 2[b]	P for trend
All ages								
Unexposed	2360	12.8 (13.0)	Reference	0.78	Reference	0.87	Reference	0.36
Moderately	1949	12.7 (12.7)	-0.15 (-0.93; 0.64)		-0.16 (-0.93; 0.62)		-0.41 (-1.15; 0.33)	
Severely	783	12.7 (14.1)	-0.09 (-1.15; 0.96)		0.20 (-0.85; 1.25)		-0.32 (-1.32; 0.67)	
0-9 years								
Unexposed	1552	13.8 (13.6)	Reference	0.78	Reference	0.89	Reference	0.31
Moderately	1179	13.6 (13.6)	-0.19 (-1.22; 0.85)		-0.29 (-1.32; 0.74)		-0.51 (-1.49; 0.46)	
Severely	447	13.7 (14.3)	-0.12 (-1.56; 1.31)		0.07 (-1.35; 1.50)		-0.53 (-1.88; 0.82)	
10-17 years								
Unexposed	808	10.9 (11.6)	Reference	0.46	Reference	0.61	Reference	0.83
Moderately	770	11.2 (11.2)	0.30 (-0.87; 1.47)		0.06 (-1.11; 1.23)		-0.21 (-1.33; 0.90)	
Severely	336	11.5 (13.7)	0.53 (-0.98; 2.04)		0.44 (-1.06; 1.94)		-0.09 (-1.53; 1.35)	

[a] only current drinkers (>1 g/day);
[b] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level;
multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, smoking status and intensity, and mMDS.

Supplemental table 5 Prevalence ratios and 95% CI for having an unhealthy diet[a], according to level of famine exposure, stratified by age category.

Age category and famine exposure level	Crude model	P for trend	Multivariable model 1[b]	P for trend	Multivariable model 2[b]	P for trend
All ages						
Unexposed	Reference	0.60	Reference	0.55	Reference	0.26
Moderately	0.92 (0.86; 0.98)		0.92 (0.87; 0.99)		0.92 (0.86; 0.98)	
Severely	1.02 (0.94; 1.10)		1.01 (0.93; 1.09)		0.98 (0.91; 1.07)	
0-9 years						
Unexposed	Reference	0.44	Reference	0.32	Reference	0.18
Moderately	0.89 (0.82; 0.97)		0.90 (0.82; 0.98)		0.89 (0.82; 0.97)	
Severely	1.01 (0.90; 1.13)		0.99 (0.89; 1.11)		0.97 (0.87; 1.09)	
10-17 years						
Unexposed	Reference	0.98	Reference	0.76	Reference	0.80
Moderately	0.94 (0.85; 1.04)		0.96 (0.87; 1.07)		0.95 (0.86; 1.06)	
Severely	1.02 (0.90; 1.16)		1.04 (0.92; 1.17)		1.00 (0.88; 1.13)	

[a] unhealthy diet is defined as mMDS<4 (excluding alcohol);
[b] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level;
multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, smoking status and intensity, and alcohol consumption.

Supplemental table 6 Association between famine exposure and modified Mediterranean Diet Score (excluding alcohol), stratified by age category, regression coefficients and 95% CI, n=7,525.

Age category and famine exposure level	N	mMDS, mean (SD)	Crude model	P for trend	Multivariable model 1[a]	P for trend	Multivariable model 2[a]	P for trend
All ages								
Unexposed	3450	4.0 (1.5)	Reference	0.33	Reference	0.31	Reference	0.10
Moderately	2838	4.1 (1.5)	0.10 (0.03; 0.17)		0.08 (0.01; 0.16)		0.09 (0.02; 0.17)	
Severely	1237	4.0 (1.5)	0.00 (-0.09; 0.10)		0.02 (-0.08; 0.11)		0.05 (-0.05; 0.14)	
0-9 years								
Unexposed	2122	4.0 (1.5)	Reference	0.32	Reference	0.18	Reference	0.09
Moderately	1601	4.1 (1.5)	0.13 (0.03; 0.22)		0.12 (0.03; 0.22)		0.13 (0.04; 0.23)	
Severely	662	4.0 (1.6)	0.00 (-0.13; 0.13)		0.03 (-0.10; 0.16)		0.05 (-0.07; 0.18)	
10-17 years								
Unexposed	1328	4.0 (1.5)	Reference	0.70	Reference	0.97	Reference	0.56
Moderately	1237	4.0 (1.4)	0.06 (-0.05; 0.18)		0.03 (-0.08; 0.14)		0.04 (-0.07; 0.15)	
Severely	575	4.0 (1.5)	0.01 (-0.14; 0.15)		-0.02 (-0.16; 0.13)		0.03 (-0.11; 0.17)	

[a] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level;
multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, physical activity level, smoking status and intensity, and alcohol consumption.

Supplemental table 7 Prevalence ratios and 95% CI for being physically inactive, according to level of famine exposure, stratified by age category.

Age category and famine exposure level	Crude model	P for trend	Multivariable model 1[a]	P for trend	Multivariable model 2[a]	P for trend
All ages						
Unexposed	Reference	<0.0001	Reference	0.0008	Reference	0.081
Moderately	1.29 (1.07; 1.55)		1.23 (1.03; 1.48)		1.18 (0.99; 1.42)	
Severely	1.52 (1.22; 1.89)		1.42 (1.15; 1.77)		1.32 (1.06; 1.64)	
0-9 years						
Unexposed	Reference	0.21	Reference	0.34	Reference	0.48
Moderately	1.24 (0.93; 1.65)		1.23 (0.92; 1.63)		1.19 (0.90; 1.59)	
Severely	1.20 (0.82; 1.76)		1.11 (0.76; 1.63)		1.07 (0.73; 1.56)	
10-17 years						
Unexposed	Reference	0.0017	Reference	0.0008	Reference	0.0010
Moderately	1.24 (0.98; 1.57)		1.25 (0.98; 1.58)		1.26 (0.99; 1.59)	
Severely	1.57 (1.20; 2.05)		1.63 (1.24; 2.13)		1.60 (1.22; 2.09)	

[a] multivariable model 1: adjusted for age at start of the famine (October 1, 1944) and educational level; multivariable model 2: adjusted for age at start of the famine, educational level model, BMI, energy intake, smoking status and intensity, alcohol consumption, and mMDS.

Chapter 8

General discussion

GENERAL DISCUSSION

The research presented in this thesis shows that a Mediterranean-style diet in the Netherlands is associated with the lowest burden of chronic diseases. Quality-adjusted life years were used as a measure of overall disease burden. Previously, a Mediterranean-style diet was also associated with a lower disease burden when disability-adjusted life years were used to measure disease burden (1). Both measures, QALYs and DALYs, combine a person's life expectancy with a weight for the time living with a chronic disease. The combination of morbidity and mortality information, and the inclusion of duration and severity of different diseases, provide a comprehensive measure of the population health.

In traditional prospective analyses that use disease-specific morbidity as the outcome, the time of onset of that specific disease is needed as well as vital status of all cohort members. For QALYs, or DALYs, additional information is needed on life expectancy and on time of onset of all relevant diseases. Furthermore, updated quality of life data during ageing and disease progress is needed. Although QALYs and DALYs both originate from the same concepts, they are not interchangeable (2). For both measures information on disease occurrence and mortality, life expectancy, and disease weights are needed (3). However, the measures originate from different disciplines, which results in different ways of measuring these weights. DALY weights reflect the function loss due to a disease, mostly based on medical expert values. In our studies, disease weights for DALYs were derived by medical experts, based on functional status, treatment and prognosis (the Dutch Disability Weight study (4)). On the other hand, utility weights for QALYs represent individual quality of life information during a diseased period. Utility weights are based on information from a standardized health related quality of life questionnaire, such as the EQ-5D questionnaire (5). For a cohort, ideally, you prospectively follow members for disease occurrence and vital status, and administer the EQ-5D questionnaire to gather information on health related quality of life every time a new disease occurs in a person (and regularly during follow-up). Unfortunately, such detailed information at different time points is not available in most cohort studies, so assumptions have to be made regarding the utility weights. In our study we had an additional problem since EQ-5D data were not measured in EPIC-NL. Instead, data of the Dutch National Survey of General Practice (6) were used, necessitating additional assumptions (chapter 3). A prediction model was used to derive utility weights for different diseases in the survey and these weights were applied to the EPIC-NL population. Diseases that occurred seldom, i.e. specific types of cancer, were combined.

In most studies presented in this thesis we used QALYs as a summary health measure. We investigated the associations between QALYs and lifestyle factors (chapter 3) and QALYs and dietary patterns (chapter 4). A combination of never smoking, maintaining a normal weight, being physically active and adherence to a Mediterranean-style diet was associated with 1.75 more QALYs, indicating an almost 2 years longer life in good health for people with these healthy lifestyle factors. Reporting being adherent to a healthy dietary pattern was associated with 0.17 QALYs, indicating living approximately two more months in good health. In previous studies in EPIC-NL, DALYs were used to investigate the association between lifestyle factors and dietary patterns and disease burden (1, 7). Comparable results were found: a healthy lifestyle was associated with a 2 years longer life in good health. A higher modified Mediterranean Diet Score and Healthy Diet Indicator were related to more QALYs and to fewer losses in DALYs (1), both indicating a longer life in good health. Overall, stronger associations were observed when using DALYs (table 1). This is partly explained by higher disability weights compared to utility weights, e.g. 1 year living with diabetes counts as 0.2 year lost in DALYs and a minimum QALY of 0.845 (corresponding to approximately 0.15 year lost). This could suggest that the burden of a disease predicted by medical experts is higher than the burden of a disease that is experienced by patients (8). This was also shown in the Global Burden of Disease study. In 1990, disease weights for DALYs were derived from surveys with health care professionals, while in 2010 surveys with the general population were used. Weights assigned in 2010 were lower than the weights assigned in 1990 (9).

Table 1 Associations between dietary patterns and disease burden, expressed in QALYs or DALYs in the EPIC-NL cohort (regression coefficients and 95% CI).

Dietary Pattern	QALYs	DALYs
A priori		
Modified Mediterranean Diet Score	0.17 (0.05; 0.30)	-0.37 (-0.54; -0.21)
Healthy Diet Index	0.15 (0.03; 0.27)	-0.20 (-0.35; -0.04)
Dutch Healthy Diet Index	0.06 (-0.07; 0.19)	-0.10 (-0.27; 0.05)
A posteriori		
Prudent pattern	0.07 (-0.05; 0.19)	-0.33 (-0.49; -0.16)
Western pattern	-0.07 (-0.21; 0.07)	0.01 (-0.23; 0.17)

Results presented for high vs low adherence; adapted from: (1, 12)

The choice of the population that is asked to value health states influences the disability or utility weights. The general population, health care professionals, patients or caregivers, they all have their own perspective on health (10). For QALYs often the general population is asked, because healthcare costs, that are frequently the outcome of interest, are public costs. The QALYs in our studies were based on the Dutch EQ-5D tariff, from a representative Dutch population (11). Preferably, weights are used that originate from a country-specific representative population.

QALYs and DALYs are both estimates to express disease burden in a population. They are computed assuming relations between risk factors and disease outcome are causal, and in itself are not an argument for causality. They are especially helpful in expressing overall disease burden in populations for risk factors that may have opposite effects in different diseases, as was shown for alcohol consumption (chapter 5). Moderate alcohol consumption may exhibit lower incidence of cardiovascular diseases and mortality, but any consumption may increase risk of some cancers. The use of DALYs enabled us to investigate the association of alcohol consumption with the overall disease burden, which showed that moderate alcohol consumption (5-15 g for women, 5-30 g for men) was associated with living approximately 3 months longer in good health.

When interpreting our results the healthy volunteer effect in the EPIC-NL study should be taken into account, largely due to self-selection of participants. Our study population has lower disease incidence and lower mortality rates compared to the general Dutch population (13). Participants were relatively healthy at study entry. In addition, participants with prevalent diseases at study entry were excluded in our analyses, as they might have changed their lifestyle due to their disease. At the end of the follow-up period (2007), 20% of the cohort suffered from at least one disease and only 4.5% died. Furthermore, we included the most important chronic diseases in our analysis, but other diseases related to lifestyle may have been missed. These methodological issues may have resulted in an underestimation of the true associations in our studies.

DIETARY PATTERNS AND LIFESTYLE

In our studies, being adherent to a Mediterranean-style diet is most strongly associated with lowest disease burden, i.e. with approximately two months longer life in good health. In the PREDIMED trial a Mediterranean-style diet with extra olive oil or nuts is compared with adherence to a low-fat diet among Spanish persons at high cardiovascular risk. After a 4 year follow-up period cardiovascular events decreased by 30% in the intervention group (14). Furthermore, a beneficial effect was suggested for breast cancer (15). These results, together with positive results from observational

studies, provide a good basis for nutritional advice to the population. Several dietary guidelines committees already acknowledge the importance of using dietary patterns in their advice instead of separate nutrients or food groups, e.g. the recent guidelines of the USDA are more focused on dietary patterns (16).

Next to diet, other lifestyle factors are important to maintain a good health. In this thesis, a combination of healthy lifestyle factors, defined by non-smoking, a normal BMI, a higher physical activity level and adherence to a Mediterranean-style diet, was associated with a longer life in good health.

The next challenge is persuading people to adhere to above defined healthy lifestyle. One option is: tailoring of guidelines. To be able to provide tailored lifestyle interventions, information is needed on whom to tailor to. People with low socio-economic status on average consume unhealthier dietary patterns. We studied whether associations between lifestyle factors and unhealthy diet across educational levels differed, but most associations were consistent across educational levels. Having an inactive lifestyle, being obese and no daily breakfast taking were more often also associated with having an unhealthy diet in both high and low educated persons. Differences were only found for current and former smokers: in low educated persons, current smoking was related to an unhealthy diet, whereas in high educated persons current and former smoking was related to a healthier diet. Our results suggest that clustering of lifestyle factors should be taken into account in programs that aim to improve health, both for low and high educated persons. It should be noted that, although we found consistent results across educational levels, people with a low socio-economic status still are a specific target group needing approaches that take into account their social practice and daily circumstances (17, 18). Behavioral and environmental factors are also important determinants of an unhealthy lifestyle (19). Research on understanding determinants of dietary, physical activity and sedentary behaviors is ongoing, i.e. in the DEDIPAC project that will study the determinants across the life course (20).

Another possibility to improve adherence of people to recommended dietary and physical activity guidelines is starting early in life. The basis of a healthy dietary pattern probably already starts before birth and healthy habits develop early in life. It has been postulated that the first 1,000 days in a person's life, starting at conception, influences diet later in life and can affect chronic disease risk later in life (21). This is also supported by our analyses in women exposed to famine early in their lives, although exposure to famine is an extreme situation. Exposed women had a higher prevalence of smoking and physical inactivity later in life. Additionally, dietary

behaviors during childhood have been shown to track into adulthood (22). Thus, starting with a healthy dietary pattern early in life may result in significant health gain later in life.

PUBLIC HEALTH IMPLICATIONS

We found that a combination of healthy lifestyle factors result in a longer life in good health. We also found that an unhealthy diet clustered with other unhealthy lifestyle factors, i.e. physical inactivity and obesity, both in low and high educated persons. Furthermore, we showed that early life factors are associated with lifestyle later in life. To achieve a large public health effect, promotion of a healthy lifestyle likely should start early in life.

Thus, prevention of an unhealthy lifestyle starts in childhood by teaching children about a healthy lifestyle. However, this message is challenging in the current obesogenic environment. People are encouraged more to eat unhealthily and to be physically inactive, while a healthy lifestyle should be encouraged instead. Furthermore, knowledge on what is unhealthy not automatically results in a behavior change. The best example of this is the smoking epidemic in the second half of the last century. Dietary and lifestyle behavior are determined by a multitude of factors. People are often stuck in old habits; they are unable to change their lifestyle, although they know they should (23). Transitions during life course, i.e. starting at university, having a child, moving or reaching retirement may be important target periods for lifestyle changes (24). At these moments, daily routine changes, which may help to substantially change unhealthy lifestyle habits (25). Therefore, these transitions may be interesting time points for lifestyle interventions.

In the Netherlands, a good initiative to encourage a healthy lifestyle is the Dutch community-based approach toward overweight prevention, Youth on Healthy Weight (JOGG), which started in 2010 (26). In this approach, a combination of unhealthy lifestyle factors is addressed. Its goal is to make a healthy lifestyle the 'way of life' for young people.

With respect to diet: the healthy choice should be made the easy choice, because many dietary choices are made mindlessly (27). Nudging can be used for this: adjusting the environment to lead people to more healthy choices. In a recent experiment it was shown that more healthy products were sold if they were placed at the cash register desk (28). People were still able to buy unhealthy products, but the easiest choice was to buy a healthy product. Such methods that slightly adjust the environment may be helpful to guide people into making healthier choices. In a recent book on healthy

ageing it is expressed as 'let the environment do the work'. The authors state that by adapting the environment, healthy choices are made more easily or even unconsciously (29).

In conclusion, adherence to a Mediterranean-style diet by a Dutch population is associated with the lowest future disease burden. Additionally, a combination of healthy lifestyle behaviors is related with a significant longer healthy life. Future intervention initiatives should take clustering of unhealthy lifestyle factors into account and interventions early in life likely will achieve the largest health gain.

REFERENCES

1. Struijk EA, Beulens JW, May AM, Fransen HP, Boer JM, de Wit GA, et al. Dietary patterns in relation to disease burden expressed in Disability-Adjusted Life Years. Am J Clin Nutr. 2014;100(4):1158-65.

2. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health Policy Plan. 2006;21(5):402-8.

3. Gold MR, Muennig P. Measure-dependent variation in burden of disease estimates: implications for policy. Med Care. 2002;40(3):260-6.

4. Melse JM, Essink-Bot ML, Kramers PG, Hoeymans N. A national burden of disease calculation: Dutch disability-adjusted life-years. Dutch Burden of Disease Group. Am J Public Health. 2000;90(8):1241-7.

5. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37(1):53-72.

6. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, van der ZJ. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15(1):59-65.

7. May AM, Struijk EA, Fransen HP, Onland-Moret NC, de Wit GA, Boer JM, et al. The impact of a healthy lifestyle on Disability-Adjusted Life Years: a prospective cohort study. BMC Med. 2015;13:39.

8. De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? Health Econ. 2000;9(2):109-26.

9. Salomon JA, Vos T, Hogan DR, Gagnon M, Naghavi M, Mokdad A, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. Lancet. 2012; 380(9859):2129-43.

10. Rowen D, Mulhern B, Banerjee S, Tait R, Watchurst C, Smith SC, et al. Comparison of general population, patient, and carer utility values for dementia health states. Med Decis Making. 2015;35(1):68-80.

11. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd. 2005;149(28):1574-8.

12. Fransen HP, Beulens JW, May AM, Struijk EA, Boer JMdW, G. A., Onland-Moret NC, et al. Dietary patterns in relation to quality-adjusted life years in the EPIC-NL cohort. Prev Med. 2015;Aug(77):119-24.

13. Struijk EA, May AM, Beulens JW, van Gils CH, Monninkhof EM, van der Schouw YT, et al. Mortality and cancer incidence in the EPIC-NL cohort: impact of the healthy volunteer effect. Eur J Public Health. 2015; Feb;25(1):144-9.

14. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Aros F, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med. 2013;368(14):1279-90.

15. Toledo E, Salas-Salvado J, Donat-Vargas C, Buil-Cosiales P, Estruch R, Ros E, et al. Mediterranean Diet and Invasive Breast Cancer Risk Among Women at High Cardiovascular Risk in the PREDIMED Trial: A Randomized Clinical Trial. JAMA Intern Med. 2015:1-9.

16. USDA. Scientific report of the 2015 Dietary Guidelines Advisory Committee. 2015.

17. Bukman AJ, Teuscher D, Feskens EJ, van Baak MA, Meershoek A, Renes RJ. Perceptions on healthy eating, physical activity and lifestyle advice: opportunities for adapting lifestyle interventions to individuals with low socioeconomic status. BMC Public Health. 2014;14:1036.

18. Teuscher D, Bukman AJ, van Baak MA, Feskens EJ, Renes RJ, Meershoek A. Challenges of a healthy lifestyle for socially disadvantaged people of Dutch, Moroccan and Turkish origin in the Netherlands: a focus group study. Crit Public Health. 2015;25(5):615-26.

19. Brug J. Determinants of healthy eating: motivation, abilities and environmental opportunities. Fam Pract. 2008;25 Suppl 1:i50-5.

20. Lakerveld J, van der Ploeg HP, Kroeze W, Ahrens W, Allais O, Andersen LF, et al. Towards the integration and development of a cross-European research network and infrastructure: the DEterminants of DIet and Physical ACtivity (DEDIPAC) Knowledge Hub. Int J Behav Nutr Phys Act. 2014;11:143.

21. www.thousanddays.org.

22. Craigie AM, Lake AA, Kelly SA, Adamson AJ, Mathers JC. Tracking of obesity-related behaviours from childhood to adulthood: A systematic review. Maturitas. 2011;70(3):266-84.

23. Folling IS, Solbjor M, Helvik AS. Previous experiences and emotional baggage as barriers to lifestyle change - a qualitative study of Norwegian Healthy Life Centre participants. BMC Fam Pract. 2015;16:73.

24. Engberg E, Alen M, Kukkonen-Harjula K, Peltonen JE, Tikkanen HO, Pekkarinen H. Life events and change in leisure time physical activity: a systematic review. Sports Med. 2012;42(5):433-47.

25. McNaughton SA, Crawford D, Ball K, Salmon J. Understanding determinants of nutrition, physical activity and quality of life among older adults: the Wellbeing, Eating and Exercise for a Long Life (WELL) study. Health Qual Life Outcomes. 2012;10:109.

26. Leenaars K, Jacobs-van der Bruggen M, Renders C. Determinants of successful public-private partnerships in the context of overweight prevention in Dutch youth. Prev Chronic Dis. 2013;10:E117.

27. Wansink B. From mindless eating to mindlessly eating better. Physiol Behav. 2010;100(5):454-63.

28. Kroese FM, Marchiori DR, de Ridder DT. Nudging healthy food choices: a field experiment at the train station. J Public Health 2015; Jul 17 [Epub ahead of print].

29. Westendorp R, van Bodegom D. Oud worden in de praktijk. Laat de omgeving het werk doen. Amsterdam: Atlas contact; 2015.

SUMMARY

In this thesis we investigated dietary patterns and lifestyle and their relation with disease burden. Our aim was to define the healthiest dietary pattern in the Netherlands and provide insight in the lifestyle factors that are associated with this pattern.

Dietary patterns can be defined in two ways: a priori or a posteriori. A priori patterns are based on current knowledge, i.e., based on national dietary guidelines, while a posteriori patterns are data driven. In **chapter 2** we focus on some methodological issues concerning a posteriori dietary pattern analysis. Several subjective decisions are used to determine the number of dietary patterns to retain. As different dietary pattern solutions can vary in food group composition, this may affect reported associations with disease outcome. We therefore examined reliability of different pattern solutions, 2 to 6 patterns, and applied quantitative criteria to determine the number of patterns to extract. We included the two mostly used methods for a posteriori analysis: principal component analysis (PCA) and cluster analysis. Reliability of dietary patterns differed considerably over different solutions from PCA. In contrast, cluster analysis derived generally stable, reproducible clusters across different solutions. Applied quantitative criteria were valuable for cluster analysis but not for PCA. Associations with disease risk are influenced by the number of patterns that are retained, especially when using PCA. We therefore suggested that all studies report the reasons for choosing the number of retained patterns and look at the interpretation of the patterns. For PCA, split-half techniques with CFA and visual inspection of the scree plot can be used to substantiate this choice. For cluster analysis, internal cluster validity indices can be applied to help determine the number of clusters to retain.

In the next part of this thesis we focus on lifestyle and diet. Modifiable lifestyle factors such as diet, smoking, being overweight and physical inactivity, are related to chronic disease occurrence. Their association with mortality or morbidity has been studied before, but in this thesis we relate them to overall disease burden. In **chapter 3** four modifiable lifestyle factors, smoking behavior, weight, physical activity level and diet are related to disease burden, using quality-adjusted life years (QALYs). In addition, we combine the four lifestyle factors into a healthy lifestyle score and investigate the association between this healthy lifestyle score and QALYs. QALYs are used to measure the burden of chronic diseases, which can also be seen as the healthy life expectancy. A QALY combines information on life expectancy and quality of life. Years lived with a disease are weighted and this weight reflects the reduction of quality of life that is attributable to having a specific chronic disease. To be able to calculate QALYs in our study population, the EPIC-NL cohort, we derived disease-specific utility weights in the

second Dutch National Survey of General Practice. In our study, never smoking, having a normal weight, being physically active and adherent to a Mediterranean-style diet were positively associated with healthy life expectancy. The combination of these four lifestyle factors was associated with almost 2 years of life longer in good health, 1.75 QALYs [95% CI 1.37; 2.14]. Our findings implicate that public programs aiming at improving health could benefit from targeting at a cluster of modifiable lifestyle factors.

In **chapter 4** we investigate five different dietary patterns and relate them to QALYs. We selected patterns that were most relevant for the Dutch population. Three of these patterns were defined a priori: a Mediterranean-style diet, a healthy diet according to WHO guidelines, and a Dutch healthy diet based on guidelines defined by the Dutch Health Council. Furthermore, two a posteriori derived patterns were investigated: a 'Western' and a 'prudent' pattern, based on data of the EPIC-NL cohort. A Mediterranean diet and the Healthy Diet Indicator (HDI) were associated with a longer life in good health. Participants who had a high mMDS score, indicating a Mediterranean-style diet, had 0.17 [95% CI 0.05; 0.30] more QALYs than participants with a low score, equivalent to a two months longer life in good health. Participants with a high HDI score also had more QALYs (0.15 [0.03; 0.27]) than participants with a low HDI score. These results can be used to support the nutrition guidelines.

In **chapter 5** we investigated the association between alcohol consumption and chronic disease burden, using DALYs. It is still heavily debated if alcohol consumption should be part of a healthy dietary pattern and included in public health guidelines. Moderate alcohol consumption has been associated with a reduced risk of cardiovascular disease, but an increased risk of some cancers. Summary health measures, like DALYs, can be used when a risk factor has opposite relations with different diseases, it combines these diseases and thus is able to provide the overall health consequences of alcohol consumption. We compared former drinkers, moderate current drinkers (5-14.9 g for women, 5-29.9 g for men), and heavier current drinkers (≥15 g for women, ≥30 g for men) with light current drinkers (0-4.9 g). Moderate drinkers had a lower chronic disease burden (mean DALYs: -0,27 [95% CI -0,43; -0,11]), while former drinkers had a higher disease burden (mean DALYs: 0,81 [0,03; 1,59]) than light drinkers (0-4.9 g). This association was mainly observed among older participants and was, as expected, driven by a lower disease burden due to cardiovascular diseases. Moderate alcohol consumption was associated with a lower disease burden from cardiovascular causes (mean DALYs: -0,18 [-0,29; -0,06]) but not from cancer (mean DALYs: -0,05 [-0,16; 0,06]). Extreme alcohol consumption was not associated with burden from the selected diseases. Our results are in line with previous observational studies and support the current guidelines allowing moderate alcohol consumption up to 1 drink/day for women

and up to 2 drinks/day for men. However, our results suggest that these recommendations mainly apply to middle-aged or older populations.

In chapters 6 and 7 we investigate determinants of an unhealthy diet and lifestyle. An unhealthy diet has been associated with other unhealthy lifestyle factors such as smoking and a low physical activity level before. We investigated whether these associations differ between people with a low and a high educational level in **chapter 6**. Having a Mediterranean-style diet was defined as having a healthy diet, as this diet was associated with the lowest disease burden in our study population. We performed a cross-sectional analysis in the EPIC-NL cohort. We showed that an inactive lifestyle, a large waist circumference, no dietary supplement use and skipping breakfast were associated with an unhealthy dietary pattern, both in low as well as high educated participants. Differences were only found for current and former smokers: in low educated persons, current smoking was related to an unhealthy diet in comparison to never smoking, whereas in high educated persons current and former smoking was related to a healthier diet. These results may be used in the development of targeted health promotion strategies.

In **chapter 7** we investigate if a relatively short period of transient undernutrition early in life is associated with an unhealthy lifestyle later in life. For this study, we used data of women who were exposed to the Dutch famine during childhood or adolescence. Of the 7,525 included women of EPIC-Prospect, 46% were unexposed, 38% moderately exposed and 16% severely exposed to the Dutch famine. Moderately and severely exposed women were more often former or current smokers compared to women who did not suffer from the famine: adjusted prevalence ratio 1.10 [95% CI 1.05; 1.14] and 1.18 [1.12; 1.25], respectively. They also smoked more pack years than unexposed women. Severely exposed women were more often physically inactive than unexposed women, adjusted prevalence ratio 1.32 [1.06; 1.64]. Results did not differ between exposure age categories (0-9 and 10-17 years). We found no associations of famine exposure with alcohol consumption and no dose-dependent relations with diet. This is the first study that investigated the association between a short period of extreme hunger in early life and the presence of unhealthy lifestyle factors later in life. In women, exposure to famine was associated with a higher prevalence of smoking and physical inactivity, while no clear relations were found with diet and alcohol consumption. Our results imply that, next to having direct biological effects that increase chronic disease risk, famine exposure might indirectly relate to chronic disease risk through unhealthy lifestyle factors.

In **chapter 8** we discuss differences between QALYs and DALYs and put our results in a broader perspective. We can conclude that adherence to a Mediterranean-style diet by a Dutch population is associated with the lowest future disease burden. Additionally, a combination of healthy lifestyle behaviors is related with a significantly longer healthy life.

SAMENVATTING

In dit proefschrift bestudeerden we voedingspatronen en andere leefgewoonten in relatie tot ziektelast in de Nederlandse populatie. Ons doel was inzicht te krijgen in welk voedingspatroon in Nederland tot de laagste ziektelast leidt en welke leefstijl factoren samenhangen met dit patroon.

Voedingspatronen kunnen op twee manieren bepaald worden: op basis van reeds bestaande kennis, zoals nationale voedingsrichtlijnen, worden bepaalde voedingscomponenten als gezond gedefinieerd (ook wel 'a priori' genoemd), of 'a posteriori', achteraf bepaald op basis van voedingsinname data.
In **hoofdstuk 2** richten we ons op enkele methodologische aspecten met betrekking tot de analyse van a posteriori voedingspatronen. Bij deze analyses dient de onderzoeker het aantal te extraheren patronen te kiezen. De uiteindelijke voedingspatronen kunnen daardoor in compositie en dus in hun associaties met ziekte verschillen. Wij bestudeerden de gevolgen van een verschillend aantal extracties: te weten de extractie van 2 tot 6 voedingspatronen en bestudeerden het gebruik van kwantitatieve criteria om dit aantal te bepalen. De twee meest gebruikte methoden voor het bepalen van a posteriori voedingspatronen zijn de 'principal component analyse' (PCA) en de cluster analyse. De betrouwbaarheid van patronen verschilde aanzienlijk voor de verschillende extracties bij PCA, terwijl cluster analyse over het algemeen stabiele en reproduceerbare clusters liet zien. Kwantitatieve criteria bleken waardevol te zijn voor cluster analyse, maar niet voor PCA. Associaties met ziekte werden beïnvloed door het gekozen aantal patronen, vooral bij PCA. We adviseren daarom om altijd een onderbouwing voor het aantal gekozen voedingspatronen te vermelden in studies en de interpretatie van de patronen daarin te betrekken.

In het tweede deel van dit proefschrift ligt de focus op leefstijl en voeding. Leefstijl gerelateerde factoren, zoals roken, het hebben van overgewicht en fysieke inactiviteit zijn bekende risicofactoren van chronische ziekten (o.a. hart- en vaatziekten, kanker en diabetes). In dit proefschrift bestuderen we combinaties van verschillend risicogedrag met totale ziektelast. In **hoofdstuk 3** combineren we vier leefstijlfactoren: roken, gewicht, fysieke activiteit en voeding in één leefstijlscore en bestuderen associaties met ziektelast. Om de totale ziektelast te berekenen maken we gebruik van quality-adjusted life years (QALYs). Een QALY combineert informatie over de levensverwachting van een persoon met zijn/haar kwaliteit van leven. De jaren die iemand met een ziekte leeft worden gewogen op basis van het verlies in kwaliteit van leven welke wordt veroorzaakt door de ziekte. Een QALY komt overeen met 1 jaar leven in optimale gezondheid. Voor iedere ziekte wordt een zogenaamd 'QALY-

gewicht' bepaald, dat varieert tussen 0 en 1. Een hoger aantal QALYs betekent een langer leven in goede gezondheid, als de ziektelast hoog is dan dalen de QALYs. Om QALYs te kunnen berekenen is dus informatie nodig over 1) levensverwachting, 2) datum van optreden van (relevante) ziekten en 3) ziekte-specifieke QALY-gewichten, en tenslotte 4) datum einde van ziekte (genezing, sterfte, of 'withdrawal'). In EPIC-NL is geen informatie over kwaliteit van leven bij verschillende ziektes gemeten. Wij berekenden ziekte-specifieke gewichten daarom op basis van de 'Tweede Nationale studie naar ziekten en verrichtingen in de huisartsenpraktijk' verricht door het NIVEL.
De studie liet zien dat niet roken, een normaal gewicht (BMI<25), fysiek actief zijn en een gezond voedingspatroon bij start van het EPIC-NL onderzoek geassocieerd was met een langer leven in goede gezondheid. Dit was gemiddeld bijna 2 jaar langer in goede gezondheid (1.75 QALYs [95% BI 1.37; 2.14]. Programma's die gericht zijn op het verbeteren van gezondheid dienen bij voorkeur meerdere leefgewoonten te omvatten.

In **hoofdstuk 4** bestuderen we relaties tussen voeding en ziektelast. We relateren vijf voedingspatronen aan QALYs. De vijf patronen zijn geselecteerd op basis van relevantie voor de Nederlandse bevolking. Het betreffen drie a priori voedingspatronen: 1) gebaseerd op de Nederlandse richtlijnen goede voeding (DHD-index), 2) gebaseerd op het Mediterraan voedingspatroon (gemodificeerde mediterrane dieet score (mMDS)), en 3) gebaseerd op de voedingsrichtlijnen van de Wereldgezondheidsorganisatie (WHO) (HDI). Twee a posteriori voedingspatronen werden bestudeerd: 1) een 'Westers' en 2) een 'Prudent' patroon, gebaseerd op PCA analyse in EPIC-NL.
De mMDS en de HDI zijn beide gerelateerd aan langer leven in goede gezondheid. Een hoge (vs lage) mMDS score is geassocieerd met 0.17 [95% BI 0.05; 0.30] meer QALYs. Dit komt neer op gemiddeld twee maanden langer leven in goede gezondheid. Ook een hoge HDI score is geassocieerd met meer QALYs: (0.15 [95% BI 0.03; 0.27]). Deze resultaten kunnen gebruikt worden ter onderbouwing van voedingsrichtlijnen.

In **hoofdstuk 5** bestuderen we de relatie tussen alcohol consumptie en de totale ziektelast. Hiervoor berekenen we DALYs (disability-adjusted life years) als maat voor de ziektelast. Er bestaat onzekerheid of alcohol deel uitmaakt van een gezond voedingspatroon. Matige consumptie zou het risico op hart- en vaatziekten verlagen, maar tevens het risico op sommige kankervormen verhogen. Door de relatie van alcohol met totale ziektelast te meten worden beide effecten mee genomen. We vergeleken de ziektelast van personen die gestopt waren met drinken, van matige drinkers (5-14.9 g/dag voor vrouwen, 5-29.9 g/dag voor mannen), en van zware drinkers (≥15 g/dag voor vrouwen, ≥30 g/dag voor mannen) met de ziektelast van geringe drinkers

(0-4.9 g/dag voor vrouwen en mannen). Matig drinken was geassocieerd met een lagere ziektelast (t.o.v. gering drinken, de referentie), het verschil was -0,27 DALYs [95% BI -0,43; -0,11]. Voormalig drinken was gerelateerd met een hogere ziektelast (verschil: 0,81 DALYs [95% BI 0,03; 1,59]).
De positieve relatie voor matige drinkers werd vooral gezien in oudere personen en werd, zoals verwacht, gedreven door een lagere ziektelast van hart- en vaatziekten. Excessief alcoholgebruik (≥ 30 g/dag voor vrouwen, ≥ 60 g/dag voor mannen) was niet geassocieerd met ziektelast in onze studie, het verschil met geringe drinkers was -0,02 DALYs [95% BI -0,28; 0,27]. Onze resultaten zijn vergelijkbaar met resultaten uit eerdere observationele onderzoeken. Ze onderbouwen de richtlijnen uit 2006, waarin een matige alcohol consumptie wordt geadviseerd (maximaal 1 glas per dag voor vrouwen en maximaal 2 glazen voor mannen). Echter, onze resultaten suggereren dat deze aanbevelingen vooral gelden voor personen van middelbare leeftijd en ouderen. De nieuwe richtlijnen goede voeding uit 2015 adviseren geen of maximaal 1 glas alcohol per dag voor zowel mannen als vrouwen.

In hoofdstuk 6 en 7 bestuderen we associaties tussen ongezonde voeding en andere ongezonde leefgewoonten. In **hoofdstuk 6** bestuderen we of deze relaties vergelijkbaar zijn in mensen met verschillend opleidingsniveau. Een voedingspatroon in de Mediterrane stijl werd als gezond beschouwd, omdat dit patroon gerelateerd was met de laagste ziektelast in onze studiepopulatie (zie hoofdstuk 4). Personen die niet voldoen aan dit patroon worden gezien als mensen met een ongezond voedingspatroon. Onze analyse laat zien dat een inactieve leefstijl, een hoge buikomvang, geen gebruik van voedingssupplementen en het regelmatig overslaan van ontbijt allen geassocieerd zijn met het hebben van een ongezond voedingspatroon, zowel in laag als in hoog opgeleide personen. Associaties tussen roken en ongezonde voeding verschilden echter voor hoog en laag opgeleiden. Roken was geassocieerd met een ongezond voedingspatroon bij laag opgeleiden, terwijl bij hoog opgeleide personen roken (huidig rokend of gestopt) juist geassocieerd was met een gezonder voedingspatroon. Deze resultaten kunnen gebruikt worden in de ontwikkeling van gezondheidsvoorlichting.

In **hoofdstuk 7** onderzoeken we of een korte periode van extreme honger in de kinderjaren geassocieerd is met een ongezonde leefstijl later in het leven. Hiervoor gebruiken we informatie uit de EPIC-Prospect studie. Een deel van deze studiepopulatie is blootgesteld aan de Nederlandse Hongerwinter in 1944/45. Van de 7525 geïncludeerde vrouwen was 46% niet blootgesteld, 38% matig blootgesteld en 16% ernstig blootgesteld aan de Hongerwinter. In eerdere studies bleken blootgestelde vrouwen een hoger risico te hebben op verschillende ziekten, zoals hart- en vaatziekten,

diabetes, COPD en astma. Wij richten ons in dit onderzoek op een ongezonde leefstijl. Een ongezonde leefstijl kan als intermediaire factor werken in de relatie tussen Hongerwinter blootstelling en kans op ziekten.
Matig en ernstig blootgestelde vrouwen waren vaker roker (huidig of gestopt), vergeleken met vrouwen die niet blootgesteld waren: de prevalentie ratios zijn 1.10 [95% BI 1.05; 1.14] en 1.18 [95% BI 1.12; 1.25]. Ook rookten de blootgestelde vrouwen meer (uitgedrukt in pakjaren). Ernstig blootgestelde vrouwen waren vaker fysiek inactief dan niet blootgestelde vrouwen: prevalentie ratio 1.32 [1.06; 1.64]. We hebben tevens bestudeerd of de resultaten verschilden tussen twee leeftijdsgroepen (0-9 jaar of 10-17 jaar bij blootstelling). We vonden geen verschillen. We vonden ook geen associaties tussen Hongerwinter blootstelling en alcohol consumptie en geen dosis-afhankelijke relaties met een gezond voedingspatroon. Onze resultaten impliceren dat, naast directe biologische effecten die zorgen voor een verhoogd ziekterisico, blootstelling aan extreme honger ook indirect gerelateerd kan zijn aan een verhoogd ziekterisico, door middel van een verhoogd risico op een ongezonde leefstijl.

In **hoofdstuk 8** bespreken we de verschillen tussen QALYs en DALYs en plaatsen we de resultaten in perspectief. We concluderen dat een voedingspatroon in de Mediterrane stijl gerelateerd is aan de laagste ziektelast. Daarnaast blijkt een combinatie van gezonde leefstijl factoren gerelateerd te zijn aan een langer leven in goede gezondheid. Programma's die gericht zijn op het verbeteren van gezondheid dienen daarom bij voorkeur meerdere leefgewoonten te omvatten.

DANKWOORD

Een proefschrift schrijf je niet alleen. Daarom wil ik hier graag iedereen bedanken die op een of andere manier heeft bijgedragen aan het tot stand komen van dit proefschrift: hartelijk bedankt!
De volgende mensen wil ik graag in het bijzonder bedanken:

Allereerst mijn promotor, prof. dr. P.H.M. Peeters en mijn copromotoren, dr. A.M. May en dr. ir. J.W.J. Beulens. Beste Petra, wat was ik blij toen ik in 2012 kon beginnen op het TOP project! Bedankt dat je me ook de kans hebt gegeven om de master opleiding Epidemiologie te volgen tijdens mijn promotietraject. Ik heb erg prettig met je samengewerkt de afgelopen jaren, veel van je geleerd, en met je gelachen. Fijn dat een overleg naast nuttig ook gezellig zijn!
Beste Anne en Joline: bedankt voor jullie fijne begeleiding bij mijn promotie-onderzoek. Ik vond het erg prettig dat jullie deur altijd open stond en we wekelijks een overleg hadden, al moest ik wel even slikken toen bleek dat dat overleg altijd op maandagochtend was. Het bleek echter een goede start van de week!

De leden van de beoordelingscommissie: prof. dr. G.E.H.M. Rutten, prof. dr. ir. J. Brug, prof. dr. ir. H.A. Smit, prof. dr. W.M.M. Verschuren en prof. dr. ir. E.J.M. Feskens: bedankt voor het lezen en beoordelen van mijn manuscript.

De studies die beschreven staan in dit proefschrift zijn uitgevoerd binnen het TOP project, een samenwerkingsverband tussen het Julius Centrum en het RIVM. Ardine de Wit, Charlotte Onland-Moret, Yvonne van der Schouw, Ellen Struijk, Jolanda Boer, Jeljer Hoekstra, en Bas Bueno-de-Mesquita: hartelijk dank voor jullie bijdrage aan alle artikelen! Ik vond het een zeer geslaagd project! Jolanda, jij was degene die me 4 jaar geleden attendeerde op de vrijkomende promotieplaats, nogmaals bedankt! Leuk dat we na mijn vertrek bij het RIVM alsnog samengewerkt hebben! Ook andere co-auteurs wil ik graag hartelijk danken voor hun bijdrage aan de artikelen. En niet te vergeten mijn twee stagiaires: Sharon van Straten en Annerieke de Kruik, bedankt voor jullie bijdrage aan hoofdstuk 7. Ik vond het erg leuk om jullie te begeleiden.

Voor mijn proefschrift heb ik gebruik gemaakt van data van de EPIC-NL studie. Graag wil ik alle EPIC-NL medewerkers bij het Julius Centrum en het RIVM en natuurlijk de deelnemers aan de EPIC-NL studie hartelijk bedanken!

Het was even wennen om ineens promovendus te worden na mijn jaren bij het RIVM. Dankzij de enthousiaste collega's van de onco-groep; Coby en Henk, die altijd voor alle

Julius collega's klaarstaan; en mijn mede-promovendi voelde ik me toch snel op mijn plek. Bedankt hiervoor!

Al mijn kamergenoten van kamer 5.122 en van Geuns 5.09: bedankt voor jullie gezelligheid, de leuke etentjes, feestjes en niet te vergeten de vele lunchwandelingen! Ellen, samen werkten we aan het TOP project, jij met de DALYs, ik met de QALYs. Wat was het fijn om af en toe samen te kunnen sparren in de keuken, bedankt! Jolien, Kim, Hanneke en Nini, jullie waren in het laatste jaar mijn vaste wandelmaatjes. Heerlijk om even samen uit te kunnen waaien en intussen de laatste promotie-perikelen te bespreken! Heel veel succes met de afronding van jullie proefschriften!

Mijn lieve 'Maastrichtse' vriendinnen: Martine, Marloes, Diane, Margriet, Linda, Marijke, Ylva en Loes: bedankt voor jullie interesse in al mijn proefschrift-perikelen. Eindelijk is het klaar: "Food by Heidi" ! Ik vind het heel leuk dat we, na die geweldige tijd bij Fyrfad, nog steeds veel contact hebben! Op naar het volgende etentje/sauna-dagje/weekendje!

Mijn oud-CVG-collega's: bedankt voor jullie interesse! Het is altijd leuk om weer even terug te zijn op het RIVM en bij te kletsen. Het (oud)RIVM-groepje: eindelijk is de laatste ook gepromoveerd! Fijn dat ik al mijn promotie-lief-en-leed met jullie kon delen! De laatste jaren gingen we van Basket-borrels over naar house-warmings, etentjes en kraambezoekjes. Leuk dat we contact zijn blijven houden!

Hilde, mijn creatieve nichtje, vroeger wilde jij altijd al knutselen terwijl ik liever achter de computer zat. Eindelijk komen onze kwaliteiten nu samen in dit proefschrift! Zonder jou was dit proefschrift niet zo mooi geworden, bedankt!

Mijn lieve vriendinnen, Nicky en Anouk: bedankt dat jullie er altijd voor mij zijn! Al heel wat jaren delen wij lief en leed! Anouk, waarschijnlijk ben jij een van de weinigen die niet 'in het veld' zit, maar wel precies weet waar mijn proefschrift over gaat! Heerlijk om af en toe lekker met je te sparren, jij hebt altijd goede ideeën. Jammer dat je er bij de verdediging niet bij kan zijn, maar ik weet zeker dat je aan de andere kant van de wereld aan me zal denken. Nicky, bedankt dat jij tijdens de verdediging, letterlijk, achter me staat!

Mijn schoonfamilie: Pepijn, Jan, Marieke en Marco, Eva en Marlou; en Riekie, bedankt voor jullie interesse in mijn onderzoek. Ik vind het leuk dat jullie deze mijlpaal samen met mij vieren!
Mijn nichtjes en neefje: Bente, Elin, Isen en Liva, bedankt dat Tom af en toe bij jullie mag logeren!

Mijn ouders, bedankt voor alle kansen die jullie me gegeven hebben. Een gezond en gelukkig leven start met een goede basis en die hebben jullie me gegeven! Mama, helaas heb jij mijn hele promotie traject niet meer mee mogen maken. Wat zou je trots geweest zijn! Ik mis je. Papa, ik vind het knap hoe jij de draad weer op hebt gepakt na het verlies van mama en volop geniet van het leven, samen met Riekie en al jullie kleinkinderen!

Mirjam, lieve zus, 6 jaar geleden schreef jij al in jouw proefschrift 'Op naar jouw promotie, want als er één Fransen over de dam is volgen er meer'. Je blijkt echt voorspellende gaven te hebben! Wat is het fijn dat we een beetje in hetzelfde vak zitten, bedankt voor je steun tijdens mijn promotie! Ik ben blij dat je tijdens de verdediging achter me staat!

Lieve Tom, de afgelopen jaren ben ik niet alleen gegroeid in mijn rol als onderzoeker, maar heb ik er ook een prachtige nieuwe rol bij gekregen, ik werd moeder van jou! Jij maakt mijn dag een stukje mooier. In de ochtend vertrek ik met een lach omdat ik vol enthousiasme word uitgezwaaid. En wat is er mooier dan na een dag hard werken lekker met jou te spelen, dansen, en liedjes te zingen! Ik geniet heel erg van alle mooie momenten samen met jou en weet zeker dat er nog heel veel mooie momenten volgen!

Lieve Koen, wat heb ik getwijfeld 4 jaar geleden of ik wilde gaan promoveren. Dit betekende namelijk ook dat ik mijn vaste baan op zou zeggen. Jij hebt me altijd gesteund in mijn beslissing, bedankt hiervoor! Het was soms pittig, je hebt heel wat extra 'papa-dagen' gehad, zodat ik aan mijn proefschrift kon werken. Vooral die welbekende laatste loodjes wogen zwaar, het huishouden schoot er dan soms even bij in. Bedankt dat je ook toen voor me klaar stond, en ik na een lange dag werken kon genieten van jouw kookkunsten! Ik hou van je! Laten we nu lekker gaan genieten op Ameland. X

CURRICULUM VITAE

Heidi Fransen was born in Venray, the Netherlands on March 30, 1978. After completing secondary school (Elzendaal College, Boxmeer) she studied biomedical health sciences in Nijmegen from 1996 to 2001. After her graduation she started working as a research assistant at the department of Cardiology, Academic Hospital Maastricht.

From 2004 on she worked as a researcher at the Center for Nutrition and Health, National Institute for Public Health and the Environment (RIVM). She worked on several projects in the area of nutrition and health and the Dutch food consumption surveys. In 2012 she started a PhD programme in Cancer Epidemiology at the Julius Center for Health Sciences and Primary Care at the UMC Utrecht. Her research focused on the association between dietary patterns and chronic diseases within the EPIC-NL cohort, the results are presented in this thesis. She also successfully finished the postgraduate master of Epidemiology during her PhD programme.
Heidi lives in Utrecht, together with Koen Siemerink and their son Tom (2013).

PUBLICATION LIST

Fransen HP, Beulens JW, May AM, Struijk EA, Boer JM, de Wit GA, Onland-Moret NC, van der Schouw YT, Bueno-de-Mesquita HB, Hoekstra J, Peeters PH. Dietary patterns in relation to quality-adjusted life years in the EPIC-NL cohort. Prev Med. 2015 Aug;77:119-24.

Fransen HP, May AM, Beulens JW, Struijk EA, de Wit GA, Boer JM, Onland-Moret NC, Hoekstra J, van der Schouw YT, Bueno-de-Mesquita HB, Peeters PH. Association between Lifestyle Factors and Quality-Adjusted Life Years in the EPIC-NL Cohort. PLoS One. 2014 Nov 4;9(11):e111480.

Fransen HP, May AM, Stricker MD, Boer JM, Hennig C, Rosseel Y, Ocké MC, Peeters PH, Beulens JW. A posteriori dietary patterns: how many patterns to retain? J Nutr. 2014 Aug;144(8):1274-82.

May AM, Struijk EA, **Fransen HP**, Onland-Moret NC, de Wit GA, Boer JM, van der Schouw YT, Hoekstra J, Bueno-de-Mesquita HB, Peeters PH, Beulens JW. The impact of a healthy lifestyle on Disability-Adjusted Life Years: a prospective cohort study. BMC Med. 2015 Dec;13(1):287

Struijk EA, Beulens JW, May AM, **Fransen HP**, Boer JM, de Wit GA, Onland-Moret NC, van der Schouw YT, Hoekstra J, Bueno-de-Mesquita HB, Peeters PH. Dietary patterns in relation to disease burden expressed in Disability-Adjusted Life Years. Am J Clin Nutr. 2014 Oct;100(4):1158-65.

Struijk EA, May AM, Wezenbeek NL, **Fransen HP**, Soedamah-Muthu SS, Geelen A, Boer JM, van der Schouw YT, Bueno-de-Mesquita HB, Beulens JW. Adherence to dietary guidelines and cardiovascular disease risk in the EPIC-NL cohort. Int J Cardiol. 2014 Sep 20;176(2):354-9.

Struijk EA, May AM, Beulens JW, **Fransen HP**, de Wit GA, Boer JM, Onland-Moret NC, Hoekstra J, van der Schouw YT, Bueno-de-Mesquita HB, Peeters PH. Adherence to the Dutch Guidelines for a Healthy Diet and cancer risk in the European Prospective Investigation into Cancer and Nutrition-Netherlands (EPIC-NL) cohort. Public Health Nutr. 2014 Nov;17(11):2546-53.