Cancer is a major public health problem. In the Netherlands, cancer is -after cardiovascular disease- the second major cause of mortality. Each year approximately 66,000 invasive and 6,000 non-invasive tumors are diagnosed.[1] In men, lung cancer is the most common, followed by prostate and colorectal cancer. In women breast cancer occurred most, followed by colorectal and lung cancer. Cancer is not one single disease, but the collective term for uncontrolled cell proliferation resulting in the invasion of surrounding tissue. Cancer has multiple etiologies. The causes of cancer include genetic predisposition, environmental influences, lifestyle factors (including diet) and infectious agents.

1.1 Epidemiology of diet and cancer

Generally, epidemiological and other studies suggest that diet is an important factor in the etiology of different types of cancer. Since Doll and Peto[2] estimated that diet could be responsible for one third of all cancers in the early 1980s, the epidemiological literature on diet and cancer has grown considerably.[3] Fruits and vegetables and the potentially protective compounds they contain, such as vitamins C, E, beta-carotene and folate have been hypothesized to reduce risk of several cancers although this effect does not appear to be as large as originally thought.[3;4] The types of cancer most likely to be influenced by eating fruit and vegetables are head and neck, esophageal, lung and stomach cancers. The association with bowel cancers remains equivocal. There appears to be no association between fruit and vegetables consumption and breast or prostate cancer risk. Alcohol has been consistently found to increase the risk of several types of cancer, such as mouth, esophageal, liver, breast and colorectal cancer. Meat consumption has been hypothesized to be associated with increased risk for bowel cancer, however the evidence is equivocal. Furthermore, the overall evidence does not confirm the hypothesis that higher meat consumption increases risk of several types of cancer, such as breast, lung, stomach and prostate cancer. High fat intake has been hypothesized to increase risk of breast, colorectal and prostate cancers, but cohort studies did not show a convincing association. A high consumption of milk and dairy products and a high calcium intake have been shown to be associated with a weak but consistent reduction in the risk of bowel cancer. There are indications that the consumption of milk or dairy products does not influence the risk of breast cancer and data on a possible increase of risk for prostate cancer are limited.
Thus, the effects of numerous food items and nutrients have been investigated, but to date, the reported evidence for the relation between dietary habits and a number of cancer types is not as clear-cut and consistent across studies as one would wish for causal inference.

1.2 Rationale for dietary pattern analysis

An explanation for the inconsistent evidence regarding diet and cancer could be that epidemiological studies predominantly analyze a limited number of nutrients or food items. This approach has three major limitations. First, people do not eat isolated nutrients or foods. They eat meals consisting of a variety of foods with complex combinations of numerous constituents. As the number of candidate risk factors is large, many nutrients and foods have to be studied in order to get a comprehensive overview of the possible risk factors. An analysis including a large number of foods also yields statistically significant associations on the basis of chance. Secondly, it is not possible to control for all nutrients and foods that may be confounders, as the intakes of many dietary components are highly correlated. Finally, it is quite possible that different dietary components influence disease

risk interactively. Studies that focus on single nutrients or food items do not allow to take these interactions into account. Examination of dietary patterns that allows to study the impact of many foods and their interactions simultaneously may give more insight into the relation between diet and different types of cancer than analysis of single nutrients. The first studies on dietary patterns and cancer suggested that such an approach might better explain risk of disease than a single nutrient or food approach.[5]

1.3 Statistical analysis of dietary patterns

Overall dietary exposure can be examined by a-posteriori or a-priori approaches. The a-posteriori approach is based on current knowledge about the role of foods and nutrients in the etiology of disease. The diet is scored on whether the intake of certain foods or nutrients are in compliance with nutritional recommendations which in turn are based on the current scientific evidence on the relation between diet and disease. The resulting score can be used as a dietary exposure variable. Examples of this approach are the Diet Quality Index (DQI)[6;7], and Healthy Eating Index (HEI).[8] The a-priori approach is data-driven, using statistical techniques to summarize dietary exposure. Cluster analysis aggregates persons with similar eating habits into clusters. Members of a cluster tend to be relatively similar to each other, but dissimilar to members of other clusters. In this approach a subject usually is part of only one cluster. Factor analysis groups food variables that correlate positively or negatively into a pattern. Each subject will be assigned scores that characterize the extent to which his/her diet conforms to each pattern. The outcomes of all three approaches, dietary indices, clusters and factor-scores, can be used as risk factors in the subsequent analysis to associate overall diet with the outcome of interest. Although the combination of known risk factors are likely to better explain disease than the separate risk factors, the major limitation of the dietary index approach is that it does provide little new insight. As cluster and factor analysis are data driven they are more likely to generate new hypotheses. Neither method is better than the other. Clusters are easier to interpret as they are mutually exclusive. However, intuitively, reducing all dietary information of a person into belonging to only one of a few distinct clusters that can be compared to a reference group seems to be too much loss of potentially relevant information. For the purpose of the research described in this thesis factor analysis was chosen to disentangle a subject's dietary habits into a limited number of independent dimensions that appear to be common in the population(s) under study and to study their relevance for cancer.

1.4 Objectives

The main goal of the research described in this thesis was to gain insight into whether prevailing dietary patterns *consistently* enhance or protect against cancer.

In order to achieve this goal the following questions were formulated:

(1) Is the dietary pattern approach robust?
(2) Are dietary patterns comparable across study populations?
(3) Are dietary patterns part of a broader lifestyle pattern?
(4) Are dietary patterns associated with risk of (several types of) cancer?
(5) What is the value of the dietary pattern approach, compared to the traditional epidemiological approach?

A systematic review of the literature on dietary pattern analysis through factor analysis (including 68 studies) attempted to answer mainly the first two questions. As a literature review, hampered by the huge variation in methodology employed across the studies, has only limited power to produce strong evidence, a multi-center study - the DIETSCAN project - was set up for this purpose. In this project, including four European ongoing cohort studies on diet and cancer, a common methodology was developed to conduct the factor analysis, despite the differences between the methods used to assess diet in the original cohorts. The DIETSCAN study was used in the first place to conduct sensitivity analyses to test the robustness of (subjective) factor analytic decisions on the resulting patterns. In the second place, the study was used to compare the resulting patterns across the participating cohorts. The third purpose of DIETSCAN was to test whether the associations between shared dietary patterns, if any were found, and cancer were consistent across the four cohorts.
In addition, analyses of the complete dietary patterns identified in the Netherlands Cohort Study on Diet and Cancer (NLCS) were conducted to answer the third question, i.e. the association between the patterns and other lifestyle characteristics and also the last two questions.

1.5 The cohort studies

The Netherlands Cohort Study The Netherlands Cohort Study on diet and cancer (NLCS) is a prospective cohort study, which began in September 1986. The cohort included 62,573 women and 58,279 men aged 55 to 69 years who completed a mailed, self-administered questionnaire on dietary habits and other risk factors for cancer. Because of the case-cohort design, a random subcohort of 5000 subjects was sampled from the cohort after the baseline exposure measurement. In a case-cohort approach, follow-up information on vital status of a random subcohort provides an estimate of the person-time experience of the entire cohort. Exposure data of the subcohort is combined with those of incident cases in the entire cohort.

Other cohort studies in DIETSCAN Besides the NLCS, three other cohort studies participated in the DIETSCAN project. These were The Alpha Tocopherol Beta Carotene Cancer Prevention Study (ATBC - Finland), the Swedish Mammography Cohort (SMC), and the ORmoni e Dieta nella Eziologia dei Tumori (ORDET - Italy)
The ATBC Study was a randomized placebo-controlled intervention study conducted among male smokers. The cohort consisted of 29,133 white men aged 50 to 69 years living in Finland. After the intervention ended the follow-up of cancers and deaths has been continued. The SMC study originated from population-based mammography screening in Sweden. A questionnaire that was sent together with an invitation to be screened by mammography, was returned by 66,651 women aged 40 to 76 years. The ORDET Study is a prospective cohort study on hormonal factors and diet in the etiology of breast cancer among 10,788 healthy Italian women, aged 35-69 years.
The main criteria for selecting these studies were that they were prospective cohort studies, with comprehensive dietary assessment using a validated dietary instrument and with follow-up for cancer through record linkage of the cohorts with the national or local cancer registries

1.6 Outline of the thesis

In chapter 2 the literature on dietary patterns identified by factor analysis is reviewed, and subjectivity and comparability of dietary patterns and their associations with personal and lifestyle characteristics and disease are described. Chapter 3 describes the methodology used in the Dietscan project to determine dietary patterns, the robustness of the dietary pattern approach and the major dietary patterns extracted in each of the four cohort studies. Chapter 4 illustrates that major dietary patterns are part of a broader lifestyle pattern. Associations between dietary patterns and incidence of several types of cancer are described in chapter 5 through 7. Across the four cohorts, the common dietary patterns were linked to the risk of breast[9] and colorectal cancer (chapter 5). The choice of these cancers was based on the hypothesized association with diet and a sufficient number of cases available in the cohort studies. In chapter 6 the association of dietary patterns with lung cancer in NLCS men was compared with food-based analyses. In chapter 7, the association of dietary patterns with risk of overall cancer in the NLCS is presented and compared with nutrient-based analyses. Finally, in chapter 8, the results of the research described in this thesis are discussed, the explanatory value of the dietary pattern approach is evaluated and suggestions for further research are given.

REFERENCES

1. Visser O, Coebergh JWW, Dijck JAAM, Siesling S. Incidence of cancer in the Netherlands 1998. Utrecht: Vereniging van Integrale Kankercentra (Netherlands cancer registry), 2002.

2. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J.Natl.Cancer Inst. 1981;66:1191-308.

3. Signaleringscommissie Kanker van KWF Kankerbestrijding. De rol van voeding bij het ontstaan van kanker. Amsterdam: KWF Kankerbestrijding, 2004.

4. IARC Working Group on the Evaluation of Cancer-preventive Strategies. Fruit and vegetables, IARC handbooks of cancer prevention ; 8. Lyon: IARC Press, 2003.

5. Kant AK. Indexes of overall diet quality: a review. J.Am.Diet.Assoc. 1996;96:785-91.

6. Patterson RE, Haines PS, Popkin BM. Diet quality index: capturing a multidimensional behavior. J.Am.Diet.Assoc. 1994;94:57-64.

7. Seymour JD, Calle EE, Flagg EW, Coates RJ, Ford ES, Thun MJ. Diet Quality Index as a predictor of short-term mortality in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. Am.J.Epidemiol. 2003;157:980-8.

8. Kennedy ET, Ohls J, Carlson S, Fleming K. The Healthy Eating Index: design and applications. J.Am.Diet.Assoc. 1995;95:1103-8.

9. Mannisto S, Dixon LB, Balder HF et al. Dietary patterns and the risk of breast cancer: Results from three cohort studies in the DIETSCAN project. submitted 2003.

2

Factor analysis of food consumption data: a review of the literature

Balder HF, Van den Brandt PA, Goldbohm RA

Submitted