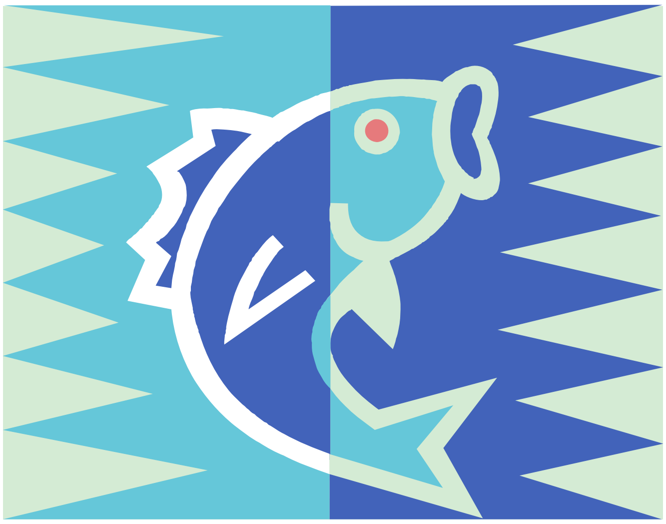
The consumption of lean and fatty fish, different dietary patterns, and the risk of cancers of various sites

Based on analyses of The Norwegian Women and Cancer (NOWAC) study and
The European Prospective Investigation into Cancer and Nutrition (EPIC) study



by
Dagrun Engeset

Institute of Community Medicine
University of Tromsø
Norway

To my family

Preface

My thesis was supposed to be about fish consumption and cancer, but when by chance I read a paper about dietary patterns my focus changed. I found it very exciting and wanted to do similar analysis on the NOWAC data. This, however, seemed to be more difficult than I had imagined since the methods used for this type of analysis was rather new within nutrition epidemiology. None of the statisticians I asked in Norway were able to assist.

However, my supervisor Eiliv Lund supported my intentions of applying these methods. And my friend and colleague, statistician Elin Alsaker, was willing to learn how to perform the analysis and assist me. Incidentally, I contacted statistician Antonio Ciampi at McGill University in Montreal, Canada, who is working with these kinds of methods. He was very enthusiastic and offered his help. Eiliv invited him to Tromsø, and the summer 2003 Ciampi spent three intensive days together with Elin and me, teaching us how to implement the analyses. This resulted in a very pleasant acquaintance, and gave fountain to my first paper. However, the original idea behind the thesis was not forgotten, and my next two analyses were on fish consumption and cancer risk.

Acknowledgment

First and foremost I would like to thank all the women participating in the Norwegian Women and Cancer study. My gratitude goes to these women who have repeatedly filled in the questionnaires and returned them to us. Many of them spent long time on the phone answering questions about their diet. Others sent us blood samples. A very impressing effort! Without their participation, this work would not have been possible to accomplish.

I am deeply grateful to my supervisor Eiliv Lund, who invited me to work with the project as a research assistant. He encouraged me to continue to work with it and gave me the opportunity to do a doctorate on both dietary patterns and fish consumption.

Statistician Elin Alsaker also deserves my deepest gratitude. Besides being a good friend, she helped me through my first two papers.

The analysis on dietary patterns had maybe not become a reality without the help from Antonio Ciampi. His enthusiasms and teaching skills has been invaluable. My first and my last paper became a reality because of him. Thank you Antonio!

I want to thank statistician Vegard Andersen who helped me with the analysis of my third paper, and also the rest of my colleagues in the NOWAC study; Marita Melhus, Bente Augdal and Guri Skeie in particular, and to everybody else who have supported me. I would also like to thank my colleagues at Departement of Biostatistics, University of Oslo, who gave me shelter when I migrated south.

I am also grateful to the Norwegian Cancer Society who supported me with a scholarship that made this project possible.

Last, but not least, a special thanks to my lovely family: husband Teddy and handsome sons William and Oliver. I dedicate this work for you all. Not because you helped me much in my work, you did rather the opposite! But you continuously reminded me that there are more important things in life - and far more challenging - than doing a doctorate: that is being a mother. The unconditional love and caring for, and from, the family!

Table of contents

Summary

The present work includes women from three cohorts: the Norwegian Women and Cancer (NOWAC) study, the European Prospective Investigation into Cancer and Nutrition (EPIC) study, and a sub-cohort from the NOWAC study termed the Norwegian EPIC cohort (NEPIC).

Despite large differences in nutritional values in lean and fatty fish most studies on fish consumption and cancer do not diverse between these two types of fish. In the present work consumption of lean and fatty fish has been examined in association with breast cancer in the EPIC study, and in association with colon cancer in the NOWAC study.

Looking at dietary patterns is a relatively new approach in nutrition epidemiology, and may give a better understanding of the relationship between diet and health. In this work, dietary patterns were identified from the NEPIC cohort, using a combination of factor and cluster analyses. The patterns were further investigated in order to reveal any relationship with cancer of various sites. Total cancer, breast cancer, and gastrointestinal cancers were examined.

There were no clear association with fish consumption and breast cancer in the EPIC study. Likewise, there were no association between fish consumption and colon cancer risk in the NOWAC study. However, when dividing lean fish into fried and poached fish, we found a statistical significant increased risk of colon cancer with high intake of poached fish.

Six different dietary patterns were identified in the NEPIC cohort. Overall, none of the dietary patterns identified was associated with cancer risk. There was, however, a somewhat higher risk of total cancer and breast cancer for women with a westernised pattern, and for total cancer for women with an alcohol pattern for some of the stratified analysis.

Samandrag (Summary in Norwegian)

Det er brukt data frå tre ulike kohortar i dette arbeidet: den norske Kvinner og kreft studien (NOWAC), The European Prospective Investigation into Cancer and Nutrition (EPIC) og ein delstudie frå Kvinner og kreftstudien kalla den norske EPIC kohorten (NEPIC).

Trass i at det er stor forskjell på næringsverdien i mager og feit fisk er det få studiar som har skilt mellom dei to fisketypane. I dette arbeidet blei det undersøkt om inntak av mager og feit fisk har nokon samanheng med brystkreft i EPIC studien og med tjukktarmkreft i NOWAC studien.

Det å undersøkje kosthaldsmønster er ei relativt ny tilnærming innan ernæringsepidemiologi og kan vere ei hjelp til å betre forstå samanhangen mellom kosthald og helse. I dette arbeidet blei det brukt ein kombinasjon av to analysemetodar, faktor- og klyngeanalyse, for å finne kosthaldsmønster i NEPIC kohorten. Deretter blei kosthaldsmønstra undersøkt for ein mogleg samanhang med ulike krefttypar. Dei ulike krefttypane som det blei undersøkt for var total kreft, brystkreft, og mage- og tarmkreft.

Det var ingen klar samanhang mellom inntak av fisk og brystkreft i EPIC studien. Det var heller ingen klar samanhang mellom fiskeinntak og tjukktarmkreft i NOWAC studien. Når mager fisk blei delt ytterlegare inn i kokt og steikt fisk såg ein likevel statistisk signifikant auka risiko for tjukktarmkreft ved høgt inntak av kokt mager fisk.

Seks ulike kosthaldsmønster blei avdekka i NEPIC kohorten. Samla sett var det ingen samanhang mellom nokon av kosthaldsmønstra og risiko for kreft. I stratifiserte analysar var der likevel ein noko høgare risiko for totalkreft og brystkreft for kvinner med vestleg kosthaldsmønster, og for totalkreft for kvinner med alkoholrelatert mønster.

List of papers

The thesis is based on the following papers, referred to in the text by their roman numerals:

I. Dietary patterns and lifestyle factors in the Norwegian EPIC cohort: The Norwegian Women and Cancer (NOWAC) study, *Eur J Clin Nutr 59(5):675-84, 2005*

II. Fish consumption and breast cancer risk. The European Prospective Investigation into Cancer and Nutrition (EPIC), *Int J Cancer 119(1):175-82, 2006*

III. Consumption of fish and risk of colon cancer in the Norwegian Women and Cancer (NOWAC) study, *Br J Nutr, 98(3): 576–582, 2007*

IV. Dietary patterns and risk of cancer of various sites in the Norwegian EPIC cohort. The Norwegian Women and Cancer (NOWAC) study, *accepted for publication in European Journal of Cancer Prevention, March 2008*

Abbreviations

CI	Confidence interval
EPIC	The European Prospective Investigation into Cancer and Nutrition
FFQ	Food frequency questionnaire
HR	Hazard ratio
NEPIC	The Norwegian EPIC cohort
NOWAC	The Norwegian Women and Cancer study
PCA	Principal Component Analysis
PH	Proportional hazard
RRR	Reduced Rank Regression
24-HDR	24-hour dietary recall
WCRF	World Cancer Research Fund

Introduction

Fish is considered to be healthy and belongs to a healthy diet. The health authorities in Norway recommends higher consumption of fish, without giving any specific amount (1). According to a recent report about fish and seafood in Norway, the average daily intake of fish in Norway is 70 grams (g); 65g for women and 75 g for men (2).

Fish consumption provides a valuable source of different nutrients like protein of high biological value, vitamin B12, selenium and iodine. Fatty fish is also a good source of vitamin D and long chain omega 3 fatty acids (1;2).

Unfortunately, fish are also prone to harmful contaminants from the environment. Known contaminants in fish are methyl mercury (3), dioxins and polychlorinated biphenyls (4). The amount and type of contaminants differ according to where the fish is caught, the fish species and methods of preparation. Dioxins and polychlorinated biphenyls accumulate in the fat, and are therefore more likely to be found in fatty fish but may also, to some extent, be present in lean fish. Methyl mercury is found in small amounts in many fish species. It accumulates in the food chain and is therefore found in the highest amount in carnivorous fishes. The amount increases with age and size of the fish (2).

Despite large differences in nutritional values in lean and fatty fish, and also in amount of contaminants, most studies on fish consumption and human health do not diverse between these two types of fish. In addition, there are studies that consider fish as white meat, and thus combine fish with chicken in the analysis. Another limitation of many fish studies is that they are performed in countries with, in average, low consumption of fish (5). There have been however, quite a few studies on long chain omega-3 fatty acids, which are found especially in fatty fish. These have been found to be beneficial for many health outcomes (6;7). Further, a Norwegian report has concluded that it will be difficult to meet the demands of vitamin D and long chain omega-3 fatty acids without fish in the diet (2). Hence the conclusion is to consume more fish.

Since fish is supposed to belong to a healthy diet, it is also interesting to look at fish in a context with other foods. One way to do this is to use several food variables from any particular survey. These variables are used as input variables in specific multivariate analysis in order to identify dietary patterns. Looking at dietary patterns is a relatively new approach in nutrition epidemiology, and may give a better understanding of the relationship between diet and health. The effect caused by specific nutrients or foods on cancer risk (or other diseases) will be diluted because of the many potential differences in nutrients between dietary patterns. Therefore, dietary pattern analysis should be considered as a complementary approach to more traditional analysis rather than a replacement (8).

Fish consumption in Norway

Norway is and has always been, a fishing nation because of its long coastline. Fish, and mainly lean white fish are consumed mostly during weekdays. The fish is eaten poached, fried, or minced and mixed with other ingredients to make fish puddings and fish cakes. The fish may be smoked, salted or marinated before cooking applies. Lean fish is considered to be daily fare. One exception is the traditional Christmas cod consumed as Christmas dinner in some districts of Norway. This cod is poached and is supposed to be as fresh as possible. Another exception is the "lutefisk". This is cod treated with lye and served poached, also for Christmas dinner.
Fatty fish, like salmon and trout, are more often used for feasts.

Some fish species, like mackerel, are considered to be inedible by people in some districts of Norway, mainly in the Northern parts of Norway. The reason is a myth that claim mackerels eat drowned sailors and fishermen.

One of the main objectives in the NOWAC study was to investigate fish consumption in relation to health matters, especially cancer. Therefore, there have been more detailed questions about fish consumption in the dietary questionnaire. Two of the questionnaire series were distributed randomly to women in northern Norway only (1996–1997) (n 13 670), due to higher fish consumption in this area. Fish referred to in the food frequency questionnaire (FFQ) is mainly salt-water fish. The women may have reported consumption of fresh water fishes in the category “other fish” but we cannot dissociate. Salmon and trout can be seen as both fresh- and salt-water fish. However, according to questionnaires put to a small sub-sample of responders from the NOWAC study, the main consumption of both salmon and trout is by farmed fish (unpublished data). The fish farms are found in the fjords, meaning salt water.

Fish consumption in Europe

The European Prospective Investigation into Cancer and Nutrition (EPIC) study revealed large differences in fish consumption among the ten countries involved in the study. Spain had the highest consumption and The Netherlands the lowest among women. Norway came as the second highest (9). The difference in the amount of fish consumed is mainly due to tradition and availability of fresh fish in the different countries and centres (9). However, the differences within the EPIC countries in fish intake are not only in the amount of fish consumed but also in the different types of fish (lean or fatty), species (cod, salmon, mullet etc.), preparation methods (frying, deep frying, poaching etc.), and how fish is consumed (with sauce, vegetables, salted etc.) (9).

In addition, participants from the other EPIC countries consumed fish mostly in weekends (Friday-Sunday), as opposed to Norway where fish is eaten mainly during weekdays (9).

There were big differences in consumption between the FFQ and the 24-hour dietary recalls (24-HDR) that were done, assuming there are an overestimation of fish consumption in the Norwegian FFQ. Still, after correcting for the overestimation, Norway remained the second highest consumer of fish.

The different fish species has not been taken into consideration in this work. Neither the differences between cold-water fish and fish from tempered water, nor differences between fresh-water fish and saltwater fish has been considered, although this is a matter that needs more attention. Fish referred to in this work is mainly saltwater fish.

Dietary patterns

Analysing dietary patterns may be a useful tool to explain the complex relationship between diet and health. We eat complex food, not single nutrients or food items; therefore the focus in nutrition research has recently turned more towards analysing dietary patterns (10-12). Identifying dietary patterns are usually done through either *a-priori* or *a-posteriori* approaches. The *a priori* approach basically means that you decide before the analysis what patterns you are looking for. Typically, you make indexes like “the Healthy Eating Index” or “the Mediterranean Diet Index” to form a diet score. In the *a-posteriori* approach the patterns are formed after doing the analysis. There are three methods belonging to this approach: factor analysis, cluster analysis and reduced rank regression (RRR). The methods take advantage of the correlations between foods. Either subjects or foods are grouped into interpretable factors or “patterns” based on the correlation structures in the dietary data. In this work we have used a combination of two of the last methods mentioned: the factor analysis and cluster analysis, but a brief description of all the methods is given in the next paragraphs.

Diet scores

With this method the data is explored using predefined combinations of foods in a diet index. The index is based on knowledge from previous studies or from deduction, and they typically rank subjects by their computed score into healthy or less healthy consumers (13). Dietary scores have been used to explore the multiple associations between the Mediterranean diet, and health parameters such as life expectancy or the incidence of obesity, cardiovascular diseases and some types of cancers. Moreover, these indexes may be useful tools to measure food consumption trends and to identify the involved factors, as well as to develop comprehensive public health nutrition recommendations.

Cluster analysis algorithms

Cluster analysis is a group of multivariate techniques used to classify *subjects* into non-overlapping groups on basis of similarity in food intake. The clusters, or groups, have high internal homogeneity (within cluster) and high external heterogeneity (between cluster) Persons in the same cluster have a food intake more comparable to each other than they have to persons in other clusters (14). The number of clusters must be given in advance, normally as the most plausible or interpretable of several tested options (13). The selection of the final cluster solution requires substantial researcher judgment, and is considered by many as too subjective (14).

Factor analysis

Factor analysis, including both principal component analysis (PCA) and common factor analysis, is a statistical approach that can be used to analyse interrelationship among a large number of variables. The technique is particularly suitable for analysing patterns of complex, multidimensional relationships (14). Factor analysis summarizes food *variables* through factor scores, which are, in a current approach, linear combinations of subsets of the original variables. This can be described as aggregating interrelated variables into possibly overlapping groups, one for each factor score. One variable may participate into the definition of more than one factor score; hence it may belong to several groups. PCA may be considered as a simple version of factor analysis, and is in fact more commonly used for identifying dietary patterns than factor analysis. In this work factor analysis will be used as a collective term, including all approaches to factor analysis.

Factor analysis can be used for the purpose to reduce the number of variables, and can be directly incorporated into other multivariate techniques (this is how factor analysis was utilized in paper I), or the factor scores can be used directly to identify different patterns; variables with higher loadings are considered more important and have greater influence on the name selected to represent the pattern (14).

Reduced rank regression (RRR)

Reduced rank regression (RRR) is a rather new method for identifying dietary patterns. The method was developed to derive dietary patterns that predict a specific response, such as a biomarker or a nutrient previously associated with chronic disease outcome (15). RRR uses information both from the study, and prior information for defining responses. RRR can be used efficiently in nutritional epidemiology by choosing disease-specific response variables and determining combinations of food intake that explain as much response variation as possible (16).

Cancer

Cancer is a generic term for a group of more than 100 diseases that can affect any part of the body, and is a leading cause of death worldwide. Breast cancer is the most common cancer among women worldwide, and the largest single cause of mortality in women (17). Incidence rates are about five times higher in industrialized countries than in less developed countries and Japan (18). Breast cancer is also the most common cancer amongst Norwegian women with 2.780 new cases in 2005, followed by colorectal cancer with 1.721 new cases (19).

Dietary factors are estimated to account for approximately 30% of cancers in industrialized countries, and are thought to be the second most important factor for the prevention of cancer. Although research to date has uncovered few definite relationships between diet and cancer risk, there is sufficient evidence that dietary factors play an important role in causing cancer (18).

Aim of the study

The aim of this study was three-partite:

- To investigate possible associations with fish consumption and cancers
- To identify different dietary patterns in Norway using a combination of cluster and factor analysis
- To investigate possible associations with cancers in the different dietary patterns

Subjects and methods

The present work includes women from three cohorts: the Norwegian Women and Cancer (NOWAC) study, the European Prospective Investigation into Cancer and Nutrition (EPIC) study, and a sub-cohort termed the Norwegian EPIC cohort (NEPIC). In its origin NEPIC is a sub-cohort from the NOWAC study that was implemented in the EPIC study in 1998/99. Hence, NEPIC forms an overlap between the NOWAC and the EPIC studies (figure 1).

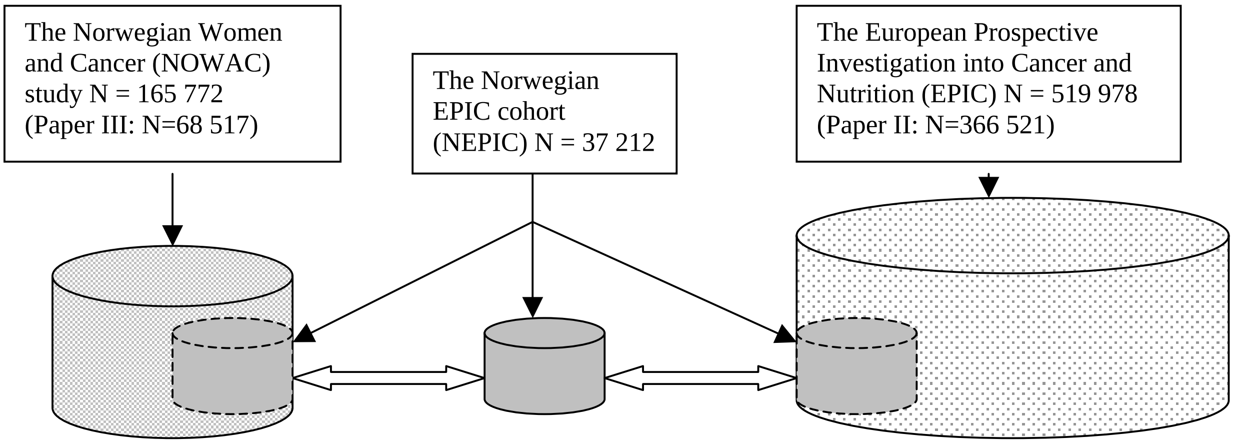


Figure 1. The different cohorts used in this work, showing what links them together.

Both cross-sectional (paper I) and prospective cohort studies (papers II-IV) are performed. The methods and study samples are described in details in the accompanying papers. Therefore, a more general description of the cohorts will be described here.

The Norwegian Women and Cancer (NOWAC) study (Paper III)

The NOWAC study was established between 1991 and 1997 as a national cohort study. The women participating in the study were selected by random sampling from the national person register, based on the unique personal number given to all inhabitants in Norway. The women were sent a self-administered questionnaire by mail where they were asked for permission to contact them again at a later stage. The women who agreed to participate have received a second questionnaire later on. Some women have also received a third questionnaire.

The questionnaires were sent in series due to restricted financial support and working capacity, and also in the interest in testing new hypothesis. Thus, the NOWAC study is composed of different sub-cohorts. In brief, NOWAC can be divided into four major sub-cohorts according to the hypothesis.

The initial questionnaires were collected between 1991-1997, the secondary questionnaires from 1998-1999, and the third mailings were sent between 2002-2005.

The questions in the different series have varied dependent on which age group they are meant for, or the hypotheses tested for, however some core questions have been common in all series. Not all series have had questions about diet (20;21).

The women have also been asked to give a blood sample. Women participating in the study currently are asked for a breast tissue sample when they visit a hospital because of breast cancer screening. The study was extended with 63 232 women in 2003-06, and at present 165 772 women participate (20). Information about blood samples or breast tissue is not utilized in this work.

The external validity of the NOWAC study has been examined, and the results showed only minor differences on factors like parity and education, and no statistical differences in other lifestyle-factors between the responders and the non-responders of the study. A postal survey among the non-responders indicated that lack of time and concern about privacy was the main reason for not responding (21).

Included in the analyses on fish consumption and colon cancer (Paper III) are 68 517 women, aged between 40-71 years (including the NEPIC cohort). They received either a first or a second questionnaire (1996 –1999), which included a semi quantitative food frequency questionnaire (FFQ) with eleven to fourteen questions about fish (Appendix A).

European Prospective Investigation into Cancer and Nutrition (EPIC) (Paper II)

EPIC is a large ongoing multi-centre prospective cohort study designed to investigate the relationship between nutrition and cancer. The study currently includes 519 978 participants (366 521 women and 153 457 men, mostly aged between 35-70 years) in 23 centres located in 10 European countries (France, Germany, Spain, the UK, the Netherlands, Denmark, Sweden, Greece, Italy and Norway). They are to be followed for cancer incidence and cause-specific mortality for several decades. During enrolment, which took place between 1992 and 2000 and at each of the different centres, information was collected through a non-dietary questionnaire on lifestyle variables and through a dietary questionnaire addressing usual diet. To calibrate dietary measurements, a standardised, computer-assisted 24-hour dietary recall (24-HDR) (EPIC- SOFT) was implemented at each centre on stratified random samples of the participants, for a total of 36

900 subjects. There have also been collected blood samples from participants in all EPIC-centres (22), but these have not been utilized in the current work.

Included in the analysis on fish consumption and breast cancer (Paper II) are women from the entire EPIC-cohort. Men are not included because the endpoint we were looking at was breast cancer, a rather rare cancer amongst men. Some centres were excluded from the analysis due to missing values for some of the adjustment variables.

The Norwegian EPIC cohort (NEPIC)(Paper I and Paper IV)

The NOWAC study was incorporated in the EPIC cohort in 1999. Women from the NOWAC study who received a second questionnaire in 1998-99, and who agreed on participating in the study, composing the Norwegian EPIC cohort (NEPIC). In total 37 226 women were included. Fourteen women have later withdrawn from the study for various reasons. The NEPIC cohort is currently consisting of 37 212 women. 1798 of these women have participated in a 24-HDR, and 12 209 women have given a blood sample.

Women who participated in the recalls and the blood sample taking were randomly selected from the entire NEPIC cohort. The recalls were carried out by phone. The women received a letter by mail asking for permission to call them and ask questions about their diet. They were further asked to inform which days and what time a day; approximately, it was best to call them. The blood samples were collected by post. The women received a small package by mail, which included the equipment needed for taking blood sample and a short questionnaire. They were asked to take this to their general practitioner if they wanted to participate. The blood sample and the filled out questionnaire were returned to us the same day from the surgery.

Only data from the main questionnaire is utilized in the analyses on dietary patterns (Papers I & IV).

Questionnaires and dietary assessments

The studies in this work are based on data from different questionnaires. The study described in Paper III is based on data from 11 questionnaire series (Figure 2), all with questions about habitual diet over the preceding year. There are different lengths on the questionnaires, the first two questionnaires had six pages and the remaining nine questionnaires had eight pages, but the FFQ's in the different series were almost similar. The reason for the different length of the questionnaires (six versus eight pages) were mainly the addition of questions about sun-bathing habits and some more questions about sweet and fatty foods in the latest questionnaires.

The study described in Paper II is based on the EPIC study and, as mentioned, data from different questionnaires from the different participating countries. The Norwegian questionnaire can be seen in the Appendix A. In addition, data from 24-HDR are used.

Paper I and IV is based on the NEPIC cohort. The questionnaire used is identical to series 28 and 29 from the NOWAC (Figure 2) and the Norwegian questionnaire in EPIC (Appendix A).

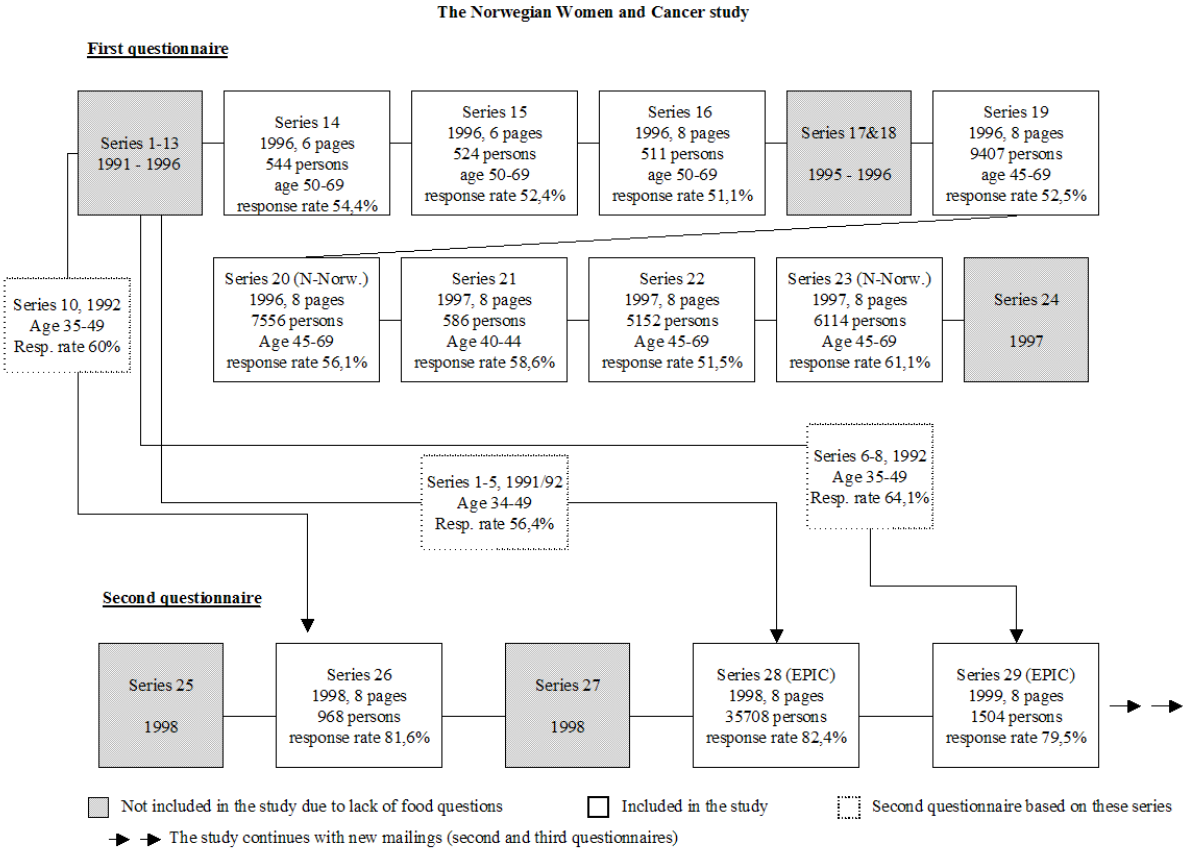


Figure 2. Flow chart of the NOWAC study, showing the series used in Paper III

Food frequency questionnaires and 24-hour dietary recalls

The first FFQ's included in the analysis in Paper III had some fewer food frequency questions than the later ones. The questions added are mostly about sweet and fatty foods, since questions about cakes and snack products were lacking in the first. Some food items have also been merged together with other similar items. For instance, in the earliest questionnaires, questions on consumption of fish balls and fish cakes were asked in two questions, whereas these were merged into one question in later questionnaires. The FFQ has been tested for validity and reproducibility.

Since the questions about fish consumption were of special interest, a validation study where the reported intake of fish was validated against the content of fatty acids in serum phospholipids was performed. The conclusion of the validation study was that the questions on fish consumption had adequate validity (23).

The FFQ has later been validated against repeated 24-HDR. The study showed that the FFQ applied in the NOWAC study performed well in estimating intake of a number of food items compared with intake estimated by repeated 24HDRs, but also that the questionnaire may overestimate the consumption of some foods and underestimate the intake of others. However, compared to FFQ's used in other large cohorts, the overall validity of the NOWAC FFQ may be described as adequate (24).

The reproducibility study concluded with a reproducibility of the dietary information within the range reported for similar instruments. However, the estimates of disease risk might be attenuated at this level of reproducibility (25).

As mentioned, the NEPIC cohort is based on two of the series described above, series 28 and 29 (Figure 2). The questionnaire is exactly the same for the two series.

In the EPIC, the FFQ's differ between countries and also between centres in the same country, both in the way they are structured and collected. In one centre, Malmö in Sweden, a non-quantitative FFQ was combined with data from a 14-day dietary diary of hot meals. The two British centres used a semi-quantitative FFQ combined with a 7-day food record. In the other centres/countries they used either an extensive quantitative FFQ or a semi-quantitative FFQ, collected either by mail or face-to-face (22). In addition, standardised 24-HDR interviews were conducted in sub-samples of participants in all the participating countries and centres, using software specially developed for this purpose, EPIC-SOFT.

EPIC-SOFT was standardised in the way that the interviews had to be conducted in the same way in all countries, but the food list and portion sizes in the program were specific to each participating country. The interviews were conducted on a random sample (8%) of the entire EPIC cohort. The 24-HDR were conducted through face-to-face interviews in all countries, except for Norway were the interviews were conducted by phone. The reason for doing the interviews by phone in Norway was because the NOWAC study is nationwide, and due to vast distances and with many sparsely populated areas, face-to-face interviews would have been very costly and almost impossible to perform on a random sample of the cohort. Foods from the FFQ were mapped to the food groups defined in EPIC- SOFT, which made direct comparison between food items from FFQ and 24-HDR possible. Data from the 24-HDR was used for calibration of data from the FFQ's to correct for systematic differences between the FFQ's.

To evaluate the validity of the method of telephone interviews versus the face-to-face approach used in the other EPIC countries, a comparison study were performed based on random sampling of two groups of women from the NEPIC.

Nutrient calculation

To be able to calculate how much the respondents had eaten in average, and the energy and nutrient level for each of the respondents, a computer program that calculated the consumption of grams eaten per day and the nutrient values of the respective food items was developed at the Institute of Community Medicine. The programme was made using SAS software, and was based on the frequency questions in the FFQ and the questions about amount of foods usually eaten.

The portion size per consumption was asked in natural or household units. A Norwegian weight and measurement table (26) was used to calculate the weights in gram for each food item. Daily intake of energy and nutrients was computed based on the Norwegian Food Composition table (27). Because of our participation in the EPIC study we had to do some work in documentation of the Norwegian Food Composition Table from 2001. Through this work we revealed a few errors in the printed version. Therefore an updated version of the table, received electronically from the responsible authorities, is used for the calculations.

Since the FFQ in series 28 and 29 was used in the EPIC study as well, another program had to be made for calculations of foods and nutrients because of the standardisation procedures in EPIC. The program was almost similar to the one used for calculations in the NOWAC study, but some adjustments had to be made. The differences in the two versions are the ability to calculate loss of nutrients with cooking of food, and fat absorption with cooking, based on coefficients from EPIC-SOFT. There are also differences based on how food items are defined; as simple food items or as recipes. As an example: we define sauce as a recipe in NOWAC, whereas in EPIC it is defined as a food, but mackerel in tomato sauce is defined as food in NOWAC and as a recipe in EPIC. The relationship between the preparations methods are estimated based on the 24-hour dietary recalls performed in Norway between 1999 and 2000, in total 1798 interviews (28). All countries participating in the EPIC study had to make documentations of their food composition tables in order to make a standardised table for analysis and calculations in EPIC (the European nutrient database, ENDB).

Non-dietary variables

All the questionnaires used in this work contained a wide range of questions about other lifestyle habits than diet. Questions about socio-economic status (e.g. income and education), physical activity, smoking habit, hormonal and reproductive factors, and anthropometrical factors like weight and height (mostly recalculated to body mass index from weight in kg and height in cm, as kg/m^2). In the NOWAC data weight and height are self-reported and are therefore prone to errors. Weight and height for most participants in the EPIC was measured at recruitment or when conducting the interviews. This makes corrections for measurement errors in the FFQ possible. We had no anthropometrical measurements taken at recruitment in NOWAC, and the 24-HDR was performed by telephone interviews, thus the measurement of weight and height was impossible. Therefore, we have no such corrections for the Norwegian weight and height.

Follow-up

For the follow-up studies described in Paper III and IV, the data collected at enrolment (1996-1999 and 1998-1999, respectively) were matched with records from the Cancer Registry of Norway to identify incident cancer cases. The data were also linked to records at Statistics Norway for information on death and emigration. The registries are based on the unique personal number system we have in Norway, and ensures an almost complete follow-up.

In Paper II, follow up is based on cancer cases reported by population based cancer registries in Denmark, the Netherlands, Spain, Sweden, the UK, and Norway, and in the Italian centres of Varese, Turin, Florence and Ragusa. Active follow-up of study participants and next-of-kin, as well as social security records and cancer and pathology registries was used in France, Germany, Greece, and the Italian centre of Naples.

The 10th Revision of the International Statistical Classification of Diseases, Injuries and Causes of Death was used for definition of cancer cases.

Statistics

Statistical analysis were conducted employing the SAS software package, version 8.02 or 9.1 for paper I, II and IV. For paper III, the statistical software program R, version 2.1.1. was used. The methods used for the different analyses are described in more details in the respective papers.

Ethics

The Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate approved the NOWAC study, and thus the NEPIC. The EPIC study was approved by the International Agency for Research on Cancer ethical committee, and by the local ethical committees in the participating countries.

Summary of results

Paper III examined fish intake and colon cancer risk in the NOWAC study. Fish intake was divided into tertiles for the analysis. We found no association between total fish intake, fatty fish intake, lean fish intake, or intake of fish products and colon cancer risk. However, when dividing lean fish into fried and poached fish, we found a statistical significant increased risk of colon cancer with high intake of poached fish (hazard ratio HR 1.46, confidence intervals CI 1.04, 2.06). Mean intake of poached lean fish in third tertile was 36.4 g/day. This equals 1.7 portions a week with a portion size of 150g. However, the association was no longer significant after exclusion of women with less than one year of follow-up (HR 1.31, CI 0.91-1.88).

Paper II examined the relationship between fish intake and breast cancer in women in the entire EPIC cohort. We stratified on pre- and post-menopausal women, and on lean and fatty fish. We found no clear association with fish consumption and breast cancer. There was a statistical significant increased risk of breast cancer in the highest quintile of fatty fish, but the test for trend was not significant. Also after stratification according to menopausal status we found no associations between fish intake and breast cancer risk in this study.

The results in Paper I revealed six different dietary patterns in the Norwegian EPIC-cohort; traditional fish consumers, healthy eaters, average group, western consumption group, traditional bread eaters, and alcohol consumption group. The different patterns had different socio-demographic and lifestyle characteristics.

In Paper IV we examined the patterns further to see if there were any association between the different patterns and cancer of different sites. We found no overall association between any patterns and the risk of breast cancer, colon cancer, rectal cancer, colorectal cancer, gastrointestinal cancer, and total cancer. When stratifying on menopausal status and smoking status, there was still no association with cancer risk. However, when stratifying on alcohol consumption, fruit and vegetable consumption, and fatty fish consumption, there was a statistical higher risk of total cancer and breast cancer with high alcohol consumption, and a significantly higher risk of breast cancer with low consumption of fruit and vegetables or with low consumption of fatty fish in the western group only. There was also a significantly higher risk of total cancer with low intake of fatty fish in the alcohol group.

General discussion

The aim of the study was tripartite, and the first part was to investigate any association between fish consumption and breast- and colon cancer. The second part of the aim was to identify dietary patterns in the NEPIC cohort. We identified six different patterns, which were used for the third part of the aim: finding associations between the patterns and cancer risk.

The strengths and limitations of the different studies that led to the findings described in this thesis have been discussed in the accompanying papers (papers I-IV). Therefore, more general issues are discussed in this chapter.

Methodological considerations in nutritional epidemiology

Epidemiological studies can be classified as experimental or observational studies (29). All studies described here are observational studies, but the study where dietary patterns were identified (paper I) has a cross-sectional design. The other three studies (paper II, III, and IV) are prospective cohort studies. In a cross-sectional study, all the information is collected at the same time, and it is therefore not possible to predict any risk of disease. In prospective cohort studies the information about a group of people are collected before exposure, and then the persons are followed for a specified period of time, normally many years, to determine the development of outcome of the study. In this work outcome was cancer of different sites.

There are many methodological aspects that need to be taken into consideration when doing analysis in nutritional epidemiology. Methodological considerations in general will be discussed in the following paragraphs. The methods used in the analysis of dietary patterns can be said to be controversial and have been criticised for being subjective, and they will therefore be discussed in a separate paragraph.

Chance

It is important in a study to try to find out whether the results were caused by chance alone. This can be assessed by hypothesis testing or by estimation and confidence intervals (CI). Hypothesis testing requires a clear statement of the hypothesis being tested and the formulation of an appropriate null hypothesis (30). A confidence interval represents the range where the variable that is being estimated is likely to be found. The effect of sample size can also be ascertained from the width of the confidence interval. A narrow confidence interval indicates less variability in the estimate of the effect and reflects a larger sample size. Wide confidence intervals indicate that the estimate is not very precise even if it is statistically significant. The results can be due to chance (31).

In the studies representing this work (papers II-IV), confidence intervals has been used as a measure of disease frequency. Paper II represent the entire female part of EPIC, except for a few centres that was excluded due to missing variables. Looking at the results for the entire cohort the confidence intervals are quite narrow, which implies that the results are most likely not due to chance alone. We found no association between fish consumption and breast cancer risk. However, looking at the risk estimates for the different countries the confidence intervals are wider. For Norway the confidence intervals were 0.36 to 5.46 in the second quintile, and 0.35 to 3.60 in the fifth quintile. Even though the risk estimates are showing null relationship with cancer, the sample size was too small (only 7 breast cancer cases in second quintile) to give enough statistical power to conclude that the results was not due to chance. There may be a possible association with breast cancer in either direction.

In paper III, representing the NOWAC study, we found a statistical significant increased risk of colon cancer with high intake of poached white fish, HR 1.46, 95% CI 1.04 to 2.06. This is a rather narrow interval and indicates that the estimate most likely is a real effect of exposure, and not due to chance. However, after exclusion of women with less than one year of follow up the risk estimate was no longer statistical significant, HR 1.31, 95% CI 0.91-1.88, but the interval is still narrow. The change in the risk estimates may indicate that the women have changed their diet due to pre-clinical symptoms of serious illness, and not that the results are due to chance. When we excluded women with less than two or three year of follow up the estimates did not change notably but the 95% CI became wider, and hence more likely to be caused by chance.

Looking at the confidence intervals in the last paper (paper IV), the intervals are quite narrow for most estimates in the main analysis, except the estimates for rectal cancer. There were few rectal cancer cases, and the estimates found here may be due to chance. The stratified analysis is not showing results for rectal cancer, or colon cancer, since the estimates was too uncertain to draw valid conclusions from.

The results from the studies presented in this work are all showing estimates with narrow confidence intervals for the main analysis, and it is therefore likely that the results are real and not due to chance.

Bias and measurement errors

Bias may be defined as any systematic error in an epidemiological study that results in an incorrect estimate of the association between exposure and risk of disease (31). A prospective cohort study is less prone to bias than a study with retrospective design where the outcome is known.

Selection bias

Any factors that affect the inclusion of participants at the beginning of a study may cause bias. The objective of drawing a sample from a population is that the sample should reflect the true effect of the population. If the sample drawn differs systematically from the population it is supposed to represent, there is a selection bias. There are different types of selection bias. Prospective studies are less open to bias than case-control studies (30), but self-selection bias or non-response bias may be a problem. In the NOWAC study, the participants have been sent a self-administered questionnaire by post. It is likely to believe that women with interest in health issues, and in cancer particular, are more prone to answer and return the questionnaire than others. Also, low response rate is considered to be a possible source of selection bias. However, an external validation study showed through different evaluations that the responders in the NOWAC study did not differ substantially from the source population, except for a somewhat higher educational level. The response rate in the NOWAC study was 57,5 % (21).

Information bias

According to Margetts and Nelson, there are three common sources of information bias (30):

1. Social desirability bias occurs when a person wants to display a desirable image, for instance when a person reports a higher intake of fruit and vegetables to appear as healthier than she really is, or overweight persons reports consuming less food and healthier food, rather than what they really eat.
2. Recall bias will always be a problem because it is difficult to remember what, and in average how much one ate the previous year. However, this is a larger problem in case- control

studies where the cases may be more likely to think thoroughly about what they did, or ate, that could have caused the disease than the controls. The role of information bias in prospective cohort studies is limited. Even if the data are biased, the bias will most likely be randomly distributed among exposed and non-exposed because information is collected before outcome of a disease. The bias may make the results of the study less distinct and can only serve to dilute any true association between the exposure and outcome.

3. Interviewer bias may occur when there is differences in the way the interviewer obtain the information. For this work this is an issue only for the analysis in paper II (the EPIC study), where the data obtained by FFQ's and dietary history was calibrated by the use of a structured computerised 24-HDR program (EPIC-SOFT). The recalls was mainly performed by face-to-face interviews by trained interviewers, with exception of Norway where we used telephone interviews. To evaluate the validity of the method of telephone interviews versus the face-to-face approach used in the other EPIC countries, a comparison study were performed based on random sampling of two groups of women from the NEPIC cohort. The telephone versus face-to-face interview design did not influence recalled diet when using the EPIC-SOFT program. However, the study confirms that 24-HDR methods are vulnerable to an interviewer effect (32).

Food frequency questionnaire (FFQ)

FFQ's are the most frequently used method for assessment of diet in cohort studies in nutritional epidemiology. They are normally not designed to assess total dietary intake but rather to record usual diet over a given period, e.g. recent months or years (semi-quantitative FFQ) (30). Data from FFQ's tend to have errors. The errors may be random or systematic, and may be caused from an inadequate food list, inappropriate response categories, or incorrect answers from the responders (see also the above paragraph on social desirability bias and recall bias). Another problem may be the time of the year the FFQ were sent. If the responders receive the questionnaire during the summer season, they may report a higher consumption, for example of cold drinks and ice cream than if they receive the questionnaire during the cold winter season.

In this work, data from different FFQ's have been used. The FFQ used for the different series in the NOWAC study are basically the same as the one used in the NEPIC cohort (Appendix A), with minor exceptions. The questionnaires used in the EPIC study differ in design and are specific for each country. The different instruments used to assess dietary intake in EPIC has been validated in the respective countries (22). As for the NOWAC study (including NEPIC), both the validity (23;24) and the reproducibility (25) of the FFQ has been examined and found acceptable.

There are, however, other aspects with a FFQ that need consideration. It is impossible to cover the entire diet; therefore it is important to think carefully what food variables to include. In NOWAC we used a semi quantitative FFQ were we tried to cover the most commonly used food items in Norway, with special emphasis on fish. For fish the seasonal variation has been taken into consideration, and for ice cream there are two separate questions: one for consumption during summer, and one for consumption during the rest of the year. It can be questioned whether there are enough variables and if the variables are appropriate to cover the common Norwegian diet. If the consumption pattern for the most commonly used food items are homogeneous, it will be difficult to find any association with the outcome variable. Then it may be necessary to include more special food items to get more variance, or heterogeneity, in the food intake. However, we have shown in the cross sectional study (Paper I) that there are heterogeneity in the food patterns: we identified six different patterns based on the food items in the questionnaire.

It may also be of interest to include more detailed questions on specific food items, like fish, that includes food preservation methods like salting, smoking or marinating, cooking methods, doneness (raw, medium or well done meat and fish), and use of condiments to get a better picture of the diet or food of interest. It may be that the questionnaires are too common to be able to reveal a true association with different diseases.

Most FFQ's used in nutrition epidemiological research cover, like the NOWAC questionnaire, only the usual or habitual diet. Very few ask more detailed questions about, for instance, the use of ready-made food versus home made food. This is a limitation with the studies. Diseases like many cancers, cardiovascular diseases, and diabetes are so-called wellness- or lifestyle diseases. These are diseases most common in rich developed countries. Use of industrialized ready-made food and sedentary work is common in rich, developed countries, whereas homemade food and a more physically active work situation are more common in poorer, less developed countries. When developing countries are changing their habits from their traditional lifestyle towards a more westernised lifestyle, there is an increase in lifestyle diseases (18;33). It might be we are missing important information by not including questions about ready made or half prepared foods in the questionnaires.

Another aspect about the food, which is difficult to cover in a FFQ, is contaminants in food derived from agriculture, packaging, food preparation, and similar.

Other issues that need consideration in a FFQ are the frequency categories and assessment of the quantities of food eaten (e.g. standard portions or household measurements). We know from NOWAC that fish consumption was over-reported (24), most likely due to more frequency choices for fish than for meat.

The data from FFQ's are often used for nutrient calculations. The calculations are based on the frequencies and quantities eaten, together with the values from a food composition table. Altogether, the values for the different nutrients must be seen as estimates with a high degree of uncertainty. Not only because of possibly incorrect answers from the responders about the frequency and quantity eaten of a food, but in addition the portion sizes are standardized and do not give the true portion for each person. Also, values in a food composition table are often based on raw, not cooked food, and many of the values are estimates, or they are average values for a limited number of samples of each food.

Due to the inaccuracy in the measurement of dietary intake in FFQ's, there have been raised concerns about the results when examining associations between diet and disease endpoints, and cancer in particular. Kristal & Potter (34) are using the term "predictive validity", and for a FFQ this is explained with the FFQ's ability to detect associations between a dietary exposure and disease outcome. Very often, when using data from a FFQ in analysis on diet and cancer, it comes out with a no-association result. For this reason many researchers have started to question the validity of FFQ's, and in at least two studies they have compared two different dietary assessment methods; a FFQ against a 7-day food record (35) and a 4-day food record (36), respectively, in studies of dietary fat and breast cancer risk. The results from the studies suggested that the predictive validity of a food record based on as few as 4 days of actual intake is superior to the predictive validity of a FFQ, at least for studies on diet and breast cancer risk

(34). Kristal & Potter speculates that cancer diseases are very complex, and thus, need more accurate dietary assessment instruments to detect associations between diet and cancer risk. They claim further that even with very large sample sizes, the error and bias in FFQ's may obscure modest associations (34).

Confounding

A confounder is defined as a factor or variable independently associated with both the exposure and the outcome variables. Confounding can lead to an over- or underestimate of the true association between exposure and outcome, and it can also change the direction of the observed effect (31). To minimize the effect of confounding, adjustment for confounding variables in multivariate analysis, like Cox Proportional Hazard (PH) analysis, or stratified analysis is important (29).

Cox (PH) analysis was used to study the association between exposure and outcome in paper II, III and IV. Potential confounders were adjusted for in all the three studies, as described in the accompanying papers. In the studies described in paper II and IV we also performed stratified analysis. Despite of this, there may be potential risk factors that we have failed to adjust for, either by ignorance or because the risk factors are presently unknown. Nevertheless, the most potential confounders are tested for, and the results should be relatively solid.

Validity

A study is considered to be valid if the findings can be considered as a reasonable representation of the true situation (30). How the data is collected has large impact on the validity of results in a study (31). The quality of the methods used in a study is often referred to as internal validity, whereas the ability to generalize the results is referred to as external validity (29). The internal validity is dependent on whether chance, bias, measurement errors and confounders are properly controlled for. We can, from the conclusions in the previous paragraphs, assume that the internal validity of the work presented here is satisfying. The external validity of the NOWAC study has been examined previously and found acceptable (21). On basis of this finding it seems reasonable to generalize the findings in this work to the female population in the same age group (40–71 y) in Norway. The external validity of the EPIC study is somewhat more difficult to interpret since the study population are samples of volunteers agreeing to participate, but not required to be random samples of defined populations. It is likely that the participants in the EPIC cohort are more health conscious than the general population. Nevertheless, any exposure-disease relationship found from the EPIC study should yield public concern.

Methodological considerations in dietary patterns

The approaches used to identify dietary patterns have been criticised for being subjective and not very reliable (37). Nevertheless, since the early 1980s the interest in dietary pattern analysis has been growing in nutritional epidemiology (12).

The analysing of dietary patterns involves many subjective choices: firstly one has to decide whether to identify the patterns by using dietary indexes, cluster analysis, factor analysis, RRR, or maybe a combination of these methods. Further, one has to decide which and how many variables to include, what form to use on the input variables (g/d, servings/day, or % energy), and whether or not to adjust for energy intake. In addition, for each method there are different options to chose from: for instance, in factor analysis one has to choose amongst several techniques of factor extraction, such as common factor analysis, PCA, and many others. Likewise, for cluster analysis there are different techniques for grouping the subjects. There is no “gold standard”, or common agreement on how to do this (12;14).

If factor analysis is the method of choice, a pattern is represented by a factor, which can be described as a dimension of variation. Thus an important problem to solve is to decide how many factors to keep. This decision should be based on several considerations: use of eigenvalue criterion (factors with eigenvalues greater than 1), a predetermined number of factors based on research objectives and/or prior research, enough factors to meet a specified percentage of variance explained, factors shown by the scree test to have substantial amounts of common variance (14). Each individual has a factor score for every derived factor (pattern) (12). In the cross sectional study (Paper I) where we identified six different dietary patterns in the NEPIC cohort, we used both the eigenvalue criterion and the scree test, and in addition we looked at the interpretability of the factors. This led to the choice of five factors, which were the factor scores used as input to the cluster analysis.

If we had used the factor scores to identify dietary patterns, we would first have to look at the initial result and decide whether to interpret the factors based on the factor loadings in each factor score. Factor scores represent the degree to which each individual score high on the group of items with high loadings on a factor. Higher values on the variables with high loadings on a factor will result in a higher factor score. Factor loadings are the correlation of each variable and the factor. Loadings indicate the degree of correspondence between the variable and the factor, with higher loadings making the variable representative of the factor. If the interpretation is difficult or do not give any meaningful information, one can decide to use a factor rotation option that can simplify the factor structure, and in most cases improves the interpretation. There are different rotation techniques to choose from. When an acceptable factor solution has been obtained, the researcher has to assign some meaning to the pattern of factor loadings. Variables with higher loadings are considered more important and have greater influence on the name or label selected to represent a factor. The final result will be a name or label that represents each of the derived factors as correct as possible (14).

Since factor scores are continuous variables and individuals have scores for each factor (pattern), it may be difficult to extrapolate from a dietary pattern derived from factor analysis to overall individual behaviour. An individual with high score in one factor may have high or low scores on others, and each individuals dietary pattern will therefore be a combination of factors, and thus not easy to interpret (12).

Cluster analysis identifies patterns in a completely different way. Instead of finding dimensions of variation, we look for homogeneous and distinct subgroups of individuals: each subgroup describes a pattern. Our choice has been to use both factor and cluster analysis in con junction. First we performed factor analysis to identify dimensions of variation. Next, without attempting a thorough interpretation of the factor scores, we used these scores as input to the cluster analysis, thus obtaining homogeneous and distinct groups of subjects. We used a two-step approach to clustering because of the large amount of data. The subjects were first clustered into 50 micro clusters using k-means clustering. This procedure groups subjects represented by the five factor scores into clusters based on Euclidian distance between observations. The 50 micro clusters were then clustered using Ward's hierarchical clustering.

The primary objective of cluster analysis is to define the structure of the data by placing the most similar observations into groups. A method that compares the similarity between the variables is required. There are several methods to choose from to measure similarity. The most commonly used measure of similarity in cluster analysis is distance measures. In our study we used the Eucleadian distance for measuring similarity, which is the most commonly recognised measure of distance. In the method we used for the final cluster solution, Ward's method for hierarchical clustering, the selection of which two clusters to combine is based on which combination of clusters minimises the within-cluster sum of squares across the complete set of separate clusters (14).

Many consider the selection of the final cluster solution as too subjective. Even though sophisticated methods have been developed to assist in evaluating the cluster solutions, it is still the researchers decision what number of clusters to choose (14). In our study, we used an index called the Calinski and Harabasz index, which is computed from the between and pooled within sum of squares. The optimal number of clusters is the numbers that maximize this index (38). We identified six clusters using this index. Whether this was the best cluster solution is difficult to say. As seen in Paper IV, a six-cluster solution made it difficult to get good estimates of cancer risks since there were few cases in each cluster or pattern. Many clusters require a longer follow up period to get enough cases in each pattern, and to get more reliable risk estimates. However, fewer clusters may hide important information that can be found with more detailed and distinct clusters. Also, a natural increase in within cluster heterogeneity comes from the reduction in number of clusters. More clusters give more within cluster homogeneity. Thus, a balance must be made defining the most appropriate number of clusters and still achieve an acceptable level of heterogeneity between the clusters (14).

Compared to factors, clusters are easier to interpret since they are mutually exclusive and continuous. An individual can belong only to one cluster (pattern) (12).

Validity and reproducibility of dietary patterns

Another limitation with the methods for identifying dietary patterns is lack of stability of the patterns and reproducibility difficulties (37). Hu et al (39) examined reproducibility and validity of dietary patterns defined by factor analysis and they concluded with reasonable reproducibility and validity of the major dietary patterns, but the patterns may not be reproducible across studies because of differences in dietary variables and the methods used for identifying the patterns. Even if there are patterns labelled "healthy" in many studies on dietary patterns, the label is a name subjectively chosen by the researcher. The variables included in the healthy patterns from different studies can vary a lot, and the differences across studies contribute to inconsistency of studies on dietary patterns.

There are very few validation studies on dietary patterns, and according to a review of the studies performed (12), only one of the validation studies considered the stability of the patterns over time, which can be seen as an indicator of reproducibility. The study compared factor solutions from a FFQ at two time points.

Internal validation of both factor and cluster analysis can be performed by splitting the study sample and repeating the analysis (12). This is what we did in Paper I to study the robustness of the patterns. We found a similar interpretation by choosing a six-cluster solution in the two random sets as in the total set. However, the six-clusters were not as clear in the two random sets as in the total set, and a four-cluster solution seemed better for the divided sets. This may be because the number of subjects in each cluster became small for some of the clusters when we split the total set, making the clusters less robust. Still, we decided to retain the six-cluster solution since all the clusters were interpretable. The six clusters had different socio-economic and lifestyle characteristics, which also gives impact to the validity of the patterns.

We also found patterns that were similar to patterns found in other studies. As mentioned, patterns with same label are not exactly like but have been identified due to the different criteria given from the researcher. All the same, the patterns had similarities both in terms of variables included in the patterns and in the characteristics of the different patterns.

Concluding remarks and future aspects

This work does not support the hypothesis about a protective effect on cancer with higher intake of fish. In general, we could not find any association between fish consumption and risk of breast- or colon cancer when looking at fish consumption alone. Neither did we find any association with fish consumption and risk of cancer in general, breast cancer, gastrointestinal cancer, colon cancer, rectal cancer, or colorectal cancer when looking at the fish pattern among the dietary patterns. However, an increased risk was seen for high intake of poached lean fish and colon cancer in the NOWAC study, and there was also an increased risk of breast cancer with high intake of fatty fish in the EPIC study, but the test for trend was not significant in the latter.

Overall, none of the dietary patterns identified in the NEPIC cohort was associated with cancer risk. We found, however, a somewhat higher risk of total cancer and breast cancer in the western group, and for total cancer in the alcohol group for some of the stratified analysis.

The findings in the studies presented here is in accordance with the recent report from World Cancer Research Fund (WCRF) who concluded with only limited evidence for a decreased risk of colorectal cancers with consumption of fish, and limited evidence for an association with fish on breast cancer risk (33), and with a Norwegian report on fish and seafood which found that the evidence for any association between fish consumption and cancer was inconclusive (2). As for dietary patterns, WCRF could give no judgement on any possible relationship between dietary patterns and risk of cancer since the existing studies were too different from each other, using different definitions for the patterns (33).

More research, preferably with different approaches, will be necessary to give any health advices since we do not know the mechanisms behind these findings. Maybe the focus in nutritional epidemiology need to change from analysis on common foods to more specified and detailed information on food and food habits. The studies on dietary patterns need to be developed further, and the aim should be to find a more standardised method for identifying patterns to make comparison between studies easier.

The lack of an overall protective effect of fish on cancer risk should not lead to advice in reducing the intake of fish since higher intake of fish may still have advantages on other diseases, like heart disease and diabetes. The increased risk we found was with high consumption of poached lean fish only, and the result needs to be confirmed by other studies.

Reference list

(1) Oppskrift for et sunnere kosthold. Handlingsplan for bedre kosthold i befolkningen (2007-2011) (In Norwegian). Departementene, Oslo; 2007.

(2) Alexander J, Frøyland L, Hemre G-I, Koster Jacobsen B, Lund E, Meltzer HM, et al. Et helhetssyn på fisk og annen sjømat i norsk kosthold (In Norwegian). Norwegian Scientific Committee for Food Safety; 2006.

(3) Sanzo JM, Dorronsoro M, Amiano P, et al. Estimation and validation of mercury intake associated with fish consumption in an EPIC cohort of Spain. Public Health Nutr 2001 Oct;4(5):981-8.

(4) Sjodin A, Hagmar L, Klasson-Wehler E, et al. Influence of the consumption of fatty Baltic Sea fish on plasma levels of halogenated environmental contaminants in Latvian and Swedish men. Environ Health Perspect 2000 Nov;108(11):1035-41.

(5) Hjartaker A. Fish consumption and risk of breast, colorectal and prostate cancer: a critical evaluation of epidemiological studies. Scandinavian Journal of Food and Nutrition 2003;47(3):111-22.

(6) Shahidi F, Miraliakbari H. Omega-3 (n-3) fatty acids in health and disease: Part 1--cardiovascular disease and cancer. J Med Food 2004;7(4):387-401.

(7) Shahidi F, Miraliakbari H. Omega-3 fatty acids in health and disease: part 2--health effects of omega-3 fatty acids in autoimmune diseases, mental health, and gene expression. J Med Food 2005;8(2):133-48.

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

(9) Welch AA, Lund E, Amiano P, et al. Variability of fish consumption within the 10 European countries participating in the European Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr 2002 Dec;5(6B):1273-85.

(10) Hoffmann K, Schulze M, Boeing H, et al. Dietary patterns: report of an international workshop. Public Health Nutr 2000;5(1):89-90.

(11) Schulze MB, Hoffmann K, Kroke A, et al. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. British Journal of Nutrition 2003 Mar;89(3):409-18.

(12) Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev 2004 May;62(5):177-203.

(13) Togo P, Osler M, Sorensen T, et al. Food intake patterns and body mass index in observational studies. International Journal of Obesity 2001 Dec;25(12):1741-51.

(14) Hair JF. Multivariate data analysis. 6th ed. ed. Upper Saddle River, N.J. : Prentice Hall; 2006.

(15) McCann SE, McCann WE, Hong CC, et al. Dietary patterns related to glycemic index and load and risk of premenopausal and postmenopausal breast cancer in the Western New York Exposure and Breast Cancer Study. Am J Clin Nutr 2007 Aug 1;86(2):465-71.

(16) Hoffmann K, Schulze MB, Schienkiewitz A, et al. Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. Am J Epidemiol 2004 May 15;159(10):935-44.

(17) World Health Organization (WHO). Fact sheet N°297. 2006.

(18) Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation. 916 ed. Geneva: World Health Organization; 2003.

(19) Cancer Registry of Norway. Cancer in Norway 2005/Kreft i Norge 2005. 2006.

(20) Lund E, Dumeaux V, Braaten T, et al. Cohort Profile: The Norwegian Women and Cancer Study—NOWAC—Kvinner og kreft. Int J Epidemiol 2007.

(21) Lund E, Kumle M, Braaten T, et al. External validity in a population-based national prospective study--the Norwegian Women and Cancer Study (NOWAC). Cancer Causes Control 2003 Dec;14(10):1001-8.

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

(23) Hjartaker A, Lund E, Bjerve KS. Serum phospholipid fatty acid composition and habitual intake of marine foods registered by a semi-quantitative food frequency questionnaire. Eur J Clin Nutr 1997 Nov;51(11):736-42.

(24) Hjartaker A, Andersen LF, Lund E. Comparison of diet measures from a food-frequency questionnaire with measures from repeated 24-hour dietary recalls. The Norwegian Women and Cancer Study. Public Health Nutr 2007 Oct;10(10):1094-103.

(25) Parr CL, Veierod MB, Laake P, et al. Test-retest reproducibility of a food frequency questionnaire (FFQ) and estimated effects on disease risk in the Norwegian Women and Cancer Study (NOWAC). Nutr J 2006;5:4.

(26) Blaker B, Aarsland M, Landsforeningen for kosthold og helse. Mål og vekt for matvarer. (In Norwegian) Oslo: Landsforeningen for kosthold & helse; 1989.

(27) Rimestad AH, Statens næringsmiddeltilsyn, Statens råd for ernæring og fysisk aktivitet, Universitetet i Oslo, Institutt for ernæringsforskning. Matvaretabellen. 2. utg ed. Oslo: Gyldendal undervisning; 2001.

(28) Skeie G. Næringsutrekning 2001 EPIC (grønt, skjematype 12, utsending 28). 2002.
Ref Type: Unpublished Work

(29) Benestad HB, Laake P. Forskningsmetode i medisin og biofag. (In Norwegian)
2nd ed. 2005.

(30) Margetts BM, Nelson M. Design Consepts in Nutritional Epidemiology. 2nd ed. Oxford University Press; 1997.

(31) Hennekens CH, Buring JE. Epidemiology in medicine. 1st ed. ed. 1987.

(32) Brustad M, Skeie G, Braaten T, et al. Comparison of telephone vs face-to-face interviews in the assessment of dietary intake by the 24[emsp14]h recall EPIC SOFT program[mdash]the Norwegian calibration study. Eur J Clin Nutr 2003 Jan;57(1):107-13.

(33) World Cancer Research Fund /American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR, 2007.

(34) Kristal AR, Potter JD. Not the time to abandon the food frequency questionnaire: counterpoint. Cancer Epidemiol Biomarkers Prev 2006 Oct;15(10):1759-60.

(35) Bingham SA, Luben R, Welch A, et al. Are imprecise methods obscuring a relation between fat and breast cancer? Lancet 2003 Jul 19;362(9379):212-4.

(36) Freedman LS, Potischman N, Kipnis V, et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. Int J Epidemiol 2006 Aug;35(4):1011-21.

(37) Jacques PF, Tucker KL. Are dietary patterns useful for understanding the role of diet in chronic disease? Am J Clin Nutr 2001 Jan;73(1):1-2.

(38) Milligan GW, Cooper MC. An Examination of Procedures for Determining the Number of Clusters in A Data Set. Psychometrika 1985;50(2):159-79.

(39) Hu FB, Rimm E, Smith-Warner SA, et al. Reproducibility and validity of dietary patterns assessed with a food- frequency questionnaire. Am J Clin Nutr 1999 Feb;69(2):243-9.

Errata

Paper III

Result section, page 578: The last sentence “The cases also consumed higher amounts of cod liver oil and vitamin D.” should be deleted.

Table 1.
The categories for years of education should be "<11", "11-13" and ">13".
For moderate physical activity, the number of cases should be 159.

Table 2.

Mean of the following foods and nutrients, distributed on tertiles T of total fish consumption in the total cohort, and among colon cancer cases (in parentheses) should be as follows:

(Mean values and standard deviations)

	T1 n=21.305 (n=71)		T2 n=21.304 (n=73)		T3 n=21.305 (n=110)	
	Mean	SD	Mean	SD	Mean	SD
Fish liver	0,1	0,2	0,2	0,4	0,4	0,6
	(0,1	0,2)	(0,2	0,4)	(0,5	0,7)
Cod liver oil	1,6	3,0	1,9	3,2	2,1	3,4
	(1,4	2,8)	(2,4	3,8)	(2,4	3,5)
Folate (μg/day)	148,9	45,3	167,7	45,9	185,8	51,0
	(138,5	37,0)	(165,5	48,7)	(179,3	51,5)
Calcium (mg/day)	632,1	257,2	682,8	255,6	736,9	273,1
	(587,5	236,8)	(640,1	245,5)	(712,6	255,4)
Fiber (g/day)	19,1	6,1	21,0	6,0	22,8	6,5
	(18,1	5,9)	(21,1	5,9)	(21,9	6,3)
Vitamin D (μg/day)	6,6	6,7	9,0	7,2	12,5	8,2
	(6,1	6,3)	(10,0	8,3)	(13,0	8,7)