Assessment of vegetable, fruit, and antioxidant vitamin intake in cancer epidemiology

Marga Ocké

Stellingen

1. De uitgebreide voedselfrequentiemethode schat de gewoonlijke groenteconsumptie slecht. *dit proefschrift.*

2. Het concept 'random bias' in de voedingsepidemiologie weerspiegelt onze onwetendheid over factoren die aan systematische meetfouten ten grondslag liggen. Het behandelen van dit type meetfout als ruis kan tot verkeerde conclusies leiden. *dit proefschrift.*

3. Herhaalde metingen van de voedselconsumptie zijn onontbeerlijk in prospectieve studies naar de relatie tussen voeding en kanker. *dit proefschrift.*

4. Het calibreren van verschillende typen voedselconsumptiemetingen in multicentrum studies met een andere voedselconsumptiemethode als referentiemeting, corrigeert wel voor methode-specifieke maar niet voor populatie-specifieke systematische meetfouten. *dit proefschrift.*

5. Het standpunt van JS Garrow (Eur J Clin Nutr 1995; 49: 231-2), dat de 'same/opposite quartile notation' informatiever is dan een correlatie-coëfficiënt voor het rapporteren van de validiteit, is onterecht. *Burema et al; Eur J Clin Nutr 1995; 49: 932-3.*

6. Investeringen in een kwalitatief goede voedingsmiddelentabel en bijbehorende kodeerafspraken door de overheid betalen zich dubbel en dwars terug.

7. Correlaties tussen schattingen van energie-inneming en energiegebruik op basis van informatie door deelnemers zijn meestal zeer laag. Dit betekent dat met minstens één van beide typen metingen iets anders wordt gemeten dan beoogd wordt.

8. In observationele epidemiologische studies naar relaties tussen voeding en chronische ziekten zouden eerst voedselpatronen in plaats van individuele voedingsstoffen bestudeerd moeten worden.

9. Het gebruik van voedingssupplementen onder het mom 'meer is beter' is niet altijd juist; het gebruik ter compensatie van een ongezond voedselpatroon is een volstrekt onjuiste benadering.

10. De meeste mensen worden niet mooier door het gebruik van make-up.

11. De georganiseerde vrijetijdsbesteding met nadruk op sensatie en spektakel verleert deelnemers om te genieten van de stilte in de natuur en maakt dit ook steeds moeilijker voor niet-deelnemers.

Stellingen behorend bij het proefschrift 'Assessment of vegetable, fruit, and antioxidant vitamin intake in cancer epidemiology' van Marga Ocké.

Wageningen, 6 maart 1996.

Assessment of vegetable, fruit, and antioxidant vitamin intake in cancer epidemiology

Marga Ocké

Promotoren:	dr. ir. D. Kromhout, Hoogleraar Volksgezondheidsonderzoek, dr. W.A. van Staveren, Bijzonder hoogleraar in de Voeding van de Oudere Mens.
Co-promotor:	dr. H.B. Bueno de Mesquita, Coördinator Kankerepidemiologie, CCM, RIVM, Bilthoven.

Assessment of vegetable, fruit, and antioxidant vitamin intake in cancer epidemiology

Marga Clara Ocké

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen
op gezag van de rector magnificus,
dr. C.M. Karssen,
in het openbaar te verdedigen
op woensdag 6 maart 1996
des namiddags te vier uur in de Aula
van de Landbouwuniversiteit te Wageningen

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Ocké, Marga Clara

Assessment of vegetable, fruit, and antioxidant vitamin intake in cancer epidemiology / Marga Clara Ocké - [S.l.: s.n.] (Wageningen: Ponsen & Looijen)
Thesis Landbouwuniversiteit Wageningen. - With ref. - With summary in Dutch.
ISBN 90-5485-506-1
Subject headings: vegetables / fruit / vitamin / cancer epidemiology / dietary intake

Printing: Grafisch Bedrijf Ponsen & Looijen b.v. Wageningen, The Netherlands.

Financial support by the National Institute of Public Health and the Environment for the publication of this thesis is gratefully acknowledged.

Abstract - Assessment of vegetable, fruit, and antioxidant vitamin intake in cancer epidemiology

PhD Thesis. Agricultural University Wageningen, the Netherlands and the National Institute of Public Health and the Environment, Bilthoven, the Netherlands.

Marga C. Ocké

Inverse associations are consistently observed in epidemiological studies on the relations between the consumption of vegetables and fruits and different types of cancer. The strength of these associations is, however, unknown amongst others because of measurement error in data on vegetable and fruit intake. The antioxidant (pro)vitamins ß-carotene, vitamin C, and vitamin E, are three of many substances in vegetables and fruits which may be responsible for the anticarcinogenic effect. This thesis is focused on the problem of intake assessment of vegetables, fruits, and antioxidant (pro)vitamins.

In the first part of the thesis, two studies on the relationships between the consumption of vegetables, fruits, and antioxidant (pro)vitamins and the occurrence of cancer are described. In the Seven Countries Study intake of vitamin C was inversely related to stomach cancer mortality at ecological level. Subjects with low intakes of vegetables, fruits, and ß-carotene that were stable over time experienced more than two-fold increased risks of lung cancer in the Zutphen Study than subjects with high stable intakes. A lack of information on the extent of measurement error in the dietary data in both studies hampered the correct interpretation of the results.

The second part of the thesis includes several studies on the estimation of measurement error in data on vegetable, fruit, and antioxidant (pro)vitamin intake and biochemical markers. In a study on the effects of frozen storage at -20 °C it was shown that vitamin E concentrations in EDTA-plasma decreased dramatically between 6 and 12 months, whereas for ß-carotene this took place after 1 year. The use of such plasma in nested case-control or case-cohort studies would result in highly attenuated odds ratios for ß-carotene and vitamin E.

Reproducibility and relative validity for food group and nutrient intake assessed with an extensive semi-quantitative food frequency questionnaire was also investigated. The questionnaire seemed adequate for ranking subjects according to intake of most nutrients and food groups including fruits, but it did not yield such good results for vegetables, ß-carotene, vitamin C for men, and vitamin E for women. The observed correlation coefficients between questionnaire and repeated 24-h recall data may be either over- or underestimates of the true validity coefficients, because of unknown error structure in both types of data. Validity coefficients estimated by a triangular comparison between questionnaire, 24-h recall, and biomarker measurements will probably be overestimates of true validity coefficients.

From these studies it is concluded that measurement error in assessing vegetable, fruit, and antioxidant (pro)vitamin intake may be large, which is a handicap for epidemiological studies. Further progress lies in improvement of dietary assessment methods, and probably even more in understanding error structures and the development of analytical methods to recognize and cope with those structures.

Contents

		page
1.	Introduction	9
2.	Average intake of antioxidant (pro)vitamins and subsequent cancer mortality in the 16 cohorts of the Seven Countries Study	19
3.	Repeated measurements of vegetables, fruit, and antioxidant (pro)vitamins in relation to lung cancer (the Zutphen Study)	33
4.	Stability of (pro)vitamins during four years of storage at -20 °C. Consequences for epidemiologic research	51
5.	The Dutch EPIC food frequency questionnaire I. Description of the questionnaire, and relative validity and reproducibility for food groups	69
6.	The Dutch EPIC food frequency questionnaire II. Relative validity and reproducibility for nutrients	91
7.	Biochemical markers as an additional measurement in dietary validity studies. Application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition	109
8.	General discussion	125
Summary		153
Samenvatting		159
Dankwoord		165
Curriculum vitae		167

Chapter 1

Introduction

Diet is known to play an important role in the development of many chronic diseases, including cancer. In 1981, Doll and Peto estimated that about 35 percent of all cancer deaths in the USA could be attributed to dietary factors with a minimum of 10 and a maximum of 70 percent [1]. Ten years later, at a conference on Nutrition and Cancer in Atlanta, USA, Doll somewhat restricted the range to 20-60 percent [2]. The still wide range illustrates the uncertainty on the state of knowledge on this topic. Next to alcohol, vegetables and fruits are the dietary components for which the relationships with cancer are best supported by the literature.

Vegetables, fruit, and cancer risk

During the last 15 years many epidemiological studies have been carried out on the relationship between the intake of vegetables and fruits and cancer at several sites. The results of these studies have been systematically reviewed in a number of papers [3-5], and the authors conclude that inverse associations are consistently observed for epithelial cancers of the respiratory and alimentary tracts. For cancers at reproductive sites fewer studies have been conducted and results are less consistent, but tend to be in the same direction for cancers of the breast, cervix, and ovary [4]. The large proportion of studies that observed inverse associations has resulted in public health recommendations to increase intake of vegetables and fruits in Europe and the United States [6,7].

The mechanism responsible for the anticarcinogenic effect of vegetables and fruits is still unclear. There are many substances in vegetables and fruits, for which protective effects have been postulated or shown experimentally. These include carotenoids, vitamin C, vitamin E, selenium, dietary fibre, dithiolthiones, glucosinolates, indoles, isothiocyanates, flavonoids, phenols, protease inhibitors, plant sterols, allium compounds, and limonene [8]. It has also been hypothesized that a mixture of compounds, which act together, instead of a single compound is responsible for the inhibition of cancer [8]. An alternative explanation would be that associations between vegetable and fruit intake and cancer risk are not causal, but that other characteristics of individuals that eat different amounts of vegetables and fruits may explain the associations. For example, a high consumption of vegetables and fruits is often associated with a low consumption of alcoholic beverages, and with less smoking in Western societies [9-11]. However, in non-Western societies dietary and lifestyle factors cluster differently and yet in those populations inverse associations between

vegetables, fruits and cancer risks have been reported too. It is therefore unlikely that one or more correlates of fruit and vegetable consumption could explain identical findings in all the different research populations [4].

The size of the reported relative risks on cancer for a low versus high intake of vegetables and fruits varies considerably. Moreover, the relative risks observed in recently published cohort studies [e.g. 12-14] are often weaker than those in case-control studies. As cohort studies are stronger by design than case-control studies, this suggests that recall, selection, or publication bias may have occurred in the case-control studies. An alternative explanation refers to the problem of random measurement error which attenuates true associations. This type of error may be larger in the recent large-scale cohort studies where time and effort that can be given to each individual for dietary intake assessment is limited.

The problem of measurement error in exposure is the central theme of this thesis. Exposures of interest are the consumption of vegetables and fruits and intake or blood levels of the antioxidant (pro)vitamins ß-carotene, vitamin C and vitamin E. Vitamin C and ß-carotene are mainly found in fruits and vegetables, whereas the main sources of vitamin E are (products rich in) vegetable oils, vegetables and cereals. These three (pro)vitamins are the most often studied components of vegetables and fruits in relation to cancer risk and datasets with information on these compounds were available to investigate measurement error in exposure and the relationship with cancer.

Assessment of vegetable, fruit, and vitamin intake

In observational studies, assessment of vegetable and fruit intake is commonly done by food consumption methods, although a search for biochemical indicators of their intake has started [15]. In order to estimate dietary intakes of ß-carotene, vitamin C, and vitamin E, food consumption methods as well as concentrations in plasma or serum are widely used, whereas concentrations of ß-carotene and vitamin E in adipose tissue have recently been added as another approach [16,17]. Similar to dietary intake assessment in general, all estimates of vegetable, fruit and antioxidant (pro)vitamin intake include inherent measurement error.

For food consumption methods which rely on information given by study participants, errors may be caused by factors related to the subjects studied, to the method, and to the processing of the data [18]. The respondent may have a memory problem in recalling what was eaten and in which amount, or in averaging

consumption frequencies or portion sizes over a longer period. This may especially be so if the diet does not follow a stable pattern. For example, the average consumption frequency of individual vegetables seems difficult to estimate because of the large number of vegetables which are each eaten infrequently. The estimation of actual or habitual portion sizes of vegetables is not easy either. For the Western diet, the problems mentioned above seem smaller for fruits, as a large proportion of total fruit intake is contributed by few fruits, e.g. apples, oranges, and bananas in the Netherlands [19]. Moreover, the amounts of fruit consumption can often be estimated in standard units. Social desirability may also influence answers on questions about diet. As vegetables and fruits are increasingly promoted as being healthy, one might expect an overestimation because of public health programs. Further, in comparison with non-obese, obese people on average underestimate energy intake [20]. Whether this is also the case for vegetables, fruits and antioxidant (pro)vitamins is unclear.

Other errors may be caused by characteristics of the food consumption method. Examples of these are: omissions of important foods in questionnaires or interviews with predefined lists of foods, the general tendency of overestimating usual consumption frequency of individual foods, the tendency to eat less complex recipes when all ingredients have to be noted and weighed, or using dietary information about an insufficient number or a non-random selection of days to derive habitual intake [21].

Additional errors are due to the processing of data, i.e. coding errors and the use of (incorrect) standard portions and other assumptions. When transforming food intake to nutrient intake, errors in food composition tables and working with average food compositions also introduce errors [22]. For ß-carotene and vitamin E, food composition tables are of limited quality and often include many missing values [23,24]. To a lesser extent this is also the case for vitamin C. Moreover, foods may vary greatly in the content of antioxidant (pro)vitamins due to differences in varieties, growing, ripening, handling, and processing of the product [25].

The main advantage of biochemical indicators of intake is their 'objectivity', i.e. they do not rely on information given by subjects. However, other factors limit the utility of many biochemical indicators to represent the (dietary) intake of specific nutrients and even more of foods or food groups. Most important is probably the degree of homeostatic control of nutrient concentrations in body compartments. For many nutrients control mechanisms cause the increase in biological concentrations to be attenuated or leveled off with higher intake. The usefulness of biomarkers is

therefore dependent on the level and range of intake to be investigated. For the vitamins C and E, but probably not for ß-carotene, some homeostatic control of blood concentrations takes place [26].

Secondly, determinants other than intake may exist, e.g. genetic, environmental, life-style or other dietary factors. It is known, for instance, that smoking, alcohol intake, and blood levels of cholesterol and triglycerides influence blood concentrations of ß-carotene [27]. Blood concentrations of vitamin C are also influenced by smoking, and by acute and chronic infections, whereas blood vitamin E levels are determined by blood lipid concentrations [26].

Thirdly, the degree of absorption and conversion of the nutrient influences levels in biological material. The absorption of ß-carotene is known to vary greatly between and within persons, and is influenced by meal composition and the food matrix in which ß-carotene is present in the gut. Conversion of ß-carotene into retinoids depends on the vitamin A status of the body [28]. Further, contamination, instability, and reactivity may introduce errors during the phases of collection, storage and laboratory analyses of the biological material. For example, temperature is known to influence concentrations of antioxidant (pro)vitamins over time [29].

Like the reference period of food consumption data, the period over which intake is reflected, also determines the suitability of biological markers to estimate long-term intake. If this is a short period, within-subject variation and systematic effects of season and time of the day the biomarker is taken, should be accounted for. Blood levels of ß-carotene, vitamins C and E are known to reflect intake over a short period of days to weeks [26], whereas ß-carotene and vitamin E concentrations in adipose tissue reflect intake over a longer period [17].

Outline of the thesis

The present thesis addresses a number of methodological issues related to intake assessment of particularly vegetables, fruits, and antioxidant (pro)vitamins. In the first part (chapters 2-3) vegetable, fruit and vitamin intake is related to cancer risk in two studies with different designs and dietary assessment methods. In the chapters 4 to 7, the estimation of measurement error in exposure is the central theme. Finally, in chapter 8, problems in the assessment of vegetable, fruit, and vitamin intake in cancer epidemiology are discussed from a general perspective.

An example of an ecological analysis of the relationships between antioxidant (pro)vitamins and cancers of the lung, stomach, and colo-rectum is presented in chapter 2. These data are taken from the prospective Seven Countries Study, in which the variation in average intake of antioxidant (pro)vitamins and cancer mortality is large. The antioxidant (pro)vitamins were analyzed in food-equivalent composites, thus avoiding errors due to food composition tables. In chapter 3, intakes of vegetables, fruits, and antioxidant (pro)vitamins are related to lung cancer risk at an individual level using the data from the Zutphen Study. In this cohort study, dietary intake was assessed repeatedly which made it possible to study relative risks for subgroups according to average or stable intake over 10 years.

In chapter 4 the stability of blood (pro)vitamins when stored at -20 °C is reported. The consequences of using blood, stored up to 4 years at this temperature, to study the relationships between blood vitamin concentrations and the occurrence of a given disease, such as cancer, are simulated. Chapters 5 and 6 concern a food frequency questionnaire newly developed for the Dutch cohorts of the EPIC-study, i.e. the European Prospective Investigation into Cancer and Nutrition [30]. The questionnaire covers many foods including several questions on fruits and a large section on vegetables. The characteristics and development of the questionnaire along with its reproducibility and relative validity for food groups and nutrients are described. A method for estimating validity coefficients using a comparison of three different types of exposure assessments is presented in chapter 7. Examples from EPIC-validation studies are used to illustrate this so-called method of triads.

The discussion in chapter 8 concerns general aspects of the assessment of vegetables, fruits, and antioxidant (pro)vitamins and covers the consequences and estimation of measurement error associated with these assessments with special reference to multicenter cohort studies.

References

1. 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: 1191-1308
2. Doll R. The lessons of life: keynote address to the Nutrition and Cancer conference. Cancer Res 1992; 52(suppl): 2024-9
3. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer. I. Epidemiology. Cancer Causes Control 1991; 2: 325-57
4. Block G, Patterson B, Subar A. Fruit, vegetables, and cancer prevention: a review of the epidemiological evidence. Nutr Cancer 1992; 18: 1-29

5. Miller AB, Berrino F, Hill M, Pietinen P, Riboli E, Wahrendorf J. Diet in the aetiology of cancer: a review. Eur J Cancer 1994; 30A: 207-20
6. EC Cancer Experts. European Code Against Cancer. Luxembourg: Commission of the European Communities, 1995
7. National Research Council, Committee on Diet and Health, Food and Nutrition Board, and Commission on Life Sciences. Diet and Health, implications for reducing chronic disease risk. Washington DC: National Academy Press, 1990
8. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer. II. Mechanisms. Cancer Causes Control 1991; 2: 427-42
9. Subar AF, Harlan LC, Mattson ME. Food and nutrient intake differences between smokers and non-smokers in the US. Am J Publ Health 1990: 80: 1323-9
10. La Vecchia C, Negri E, Franceschi S, Parazzini F, Decarli A. Differences in dietary intake with smoking, alcohol, and education. Nutr Cancer 1992; 17: 297-304
11. Zondervan KT, Ocké MC, Smit HA, Seidell JC. Do dietary and supplementary intakes of antioxidants differ with smoking status? Int J Epidemiol, in press
12. Steinmetz KA, Potter JD, Folsom AR. Vegetables, fruit, and lung cancer in the Iowa Women's Health Study. Cancer Res 1993; 53: 536-43
13. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC. Intake of fat, meat, and fiber in relation to risk of colon cancer in men. Cancer Res 1994; 54: 2390-7
14. Steinmetz KA, Kushi LH, Bostick RM, Folsom AR, Potter JD. Vegetables, fruit, and colon cancer in the Iowa women's Health Study. Am J Epidemiol 1994; 139: 1-15
15. Campbell DR, Gross MD, Martini MC, Grandits GA, Slavin JL, Potter JD. Plasma carotenoids as biomarkers of vegetable and fruit intake. Cancer Epidemiol Biomarkers Prev 1994; 3: 493-500
16. Schäfer L, Overvad K. Subcutaneous adipose-tissue fatty acids and vitamin E in humans: relation to diet and sampling site. Am J Clin Nutr 1990; 52: 486-90
17. Kardinaal AFM, Van 't Veer P, Brants HAM, Van den Berg H, Van Schoonhoven J, Hermus RJJ. Relations between antioxidant vitamins in adipose tissue, plasma, and diet. Am J Epidemiol 1995; 141: 440-50
18. Bloemberg BPM. On the effect of measurement error in nutritional epidemiology using dietary history and food frequency methodology. Thesis. Leiden: State University, 1993
19. Van der Weijden. Groot deel bevolking eet te weinig groente en fruit. Voeding 1995; 56: 10-3
20. Heitmann BL, Lissner L. Dietary underreporting by obese individuals - is it specific or non-specific? BMJ 1995; 311: 986-9
21. Bingham SA, Nelson M, Paul AA, Haraldsdottir J, Løken EB, Van Staveren WA. Methods for data collection at an individual level. In: Cameron ME, Van Staveren WA. Manual on methodology for food consumption studies. Oxford: Oxford Medical Publications, 1988: pp 53-106
22. Willett W. Nutritional Epidemiology. New York: Oxford University Press, 1990
23. Mangels AR, Holden JM, Beecher GR, Forman MR, Lanza E. Carotenoid content of fruits and vegetables: an evaluation of analytic data. J Am Diet Assoc 1993; 93: 284-96
24. Romieu I, Stampfer MJ, Stryker WS et al., Food predictors of plasma beta-carotene and alpha-tocopherol: validation of a food frequency questionnaire. Am J Epidemiol 1990; 131: 864-76

25. Paul AA, Southgate DAT. Conversion into nutrients. In: Cameron ME, Van Staveren WA. Manual on methodology for food consumption studies. Oxford: Oxford Medical Publications, 1988: pp 121-144
26. Hunter D. Biochemical indicators of dietary intake. In: Willett W. Nutritional Epidemiology. New York: Oxford University Press, 1990, pp 143-216
27. Stryker WS, Kaplan LA, Stein EA, Stampfer MJ, Sober A, Willett WC. The relation of diet, cigarette smoking, and alcohol consumption to plasma beta-carotene and alpha-tocopherol levels. Am J Epidemiol 1988; 127: 283-96
28. Van Vliet T. Van den Berg H. ß-carotene absorption and cleavage in animals and man. A review. In: Van Vliet T. Intestinal absorption and cleavage of ß-carotene in rat, hamster and human models. Thesis. Amsterdam: University of Amsterdam, 1995, pp 5-49
29. Comstock GW, Alberg AJ, Helzlsouer KJ. Reported effects of long-term freezer storage on concentrations of retinol, ß-carotene, and α-tocopherol in serum or plasma summarized. Clin Chem 1993; 39: 1075-8
30. Riboli E. Nutrition and cancer: Background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol 1992; 3: 783-91

Chapter 2

Average intake of antioxidant (pro)vitamins and subsequent cancer mortality in the 16 cohorts of the Seven Countries Study

Marga Ocké, Daan Kromhout, Alessandro Menotti, Christ Aravanis, Henry Blackburn, Ratko Buzina, Flaminio Fidanza, Annemarie Jansen, Srecko Nedeljkovic, Aulikki Nissinen, Maija Pekkarinen and Hironori Toshima

Int J Cancer 1995; 61: 480-4.

Abstract

This ecologic study aimed to investigate whether differences in population mortality from lung, stomach and colorectal cancer among the 16 cohorts of the Seven Countries Study could be explained by differences in the average intake of antioxidant (pro)vitamins. In the 1960s, detailed dietary information was collected in small subsamples of the cohorts by the dietary record method. In 1987, food equivalent composites representing the average food intake of each cohort at baseline were collected locally and analyzed in a central laboratory. The vital status of all participants was verified after 25 years of follow-up. The average intake of vitamin C was strongly inversely related to the 25-year stomach cancer mortality ($r=-0.66$, $p=0.01$), also after adjustment for smoking and intake of salt or nitrate. The average intake of α-carotene, ß-carotene, and α-tocopherol were not independently related to mortality from lung, stomach, or colorectal cancer, nor was vitamin C related to lung and colorectal cancer.

Introduction

Cancer of the lung, stomach, and colorectum are the types of cancer with the highest mortality in males [1]. It has been suggested that high intakes of the antioxidant (pro)vitamins could protect against these types of cancer, although results from epidemiological studies on individuals are not conclusive [2-4]. In analytical studies, most consistent inverse associations are observed for vitamin C and stomach cancer [2], and for carotenoids and lung cancer [3].

The findings of no association could be explained by attenuation due to relatively large measurement errors in the assessment of antioxidant (pro)vitamin intakes of individuals compared with the relatively small between-subject variation in these intakes within populations. It is therefore interesting to know whether on a population level differences in the average intake of antioxidant (pro)vitamins could explain differences in mortality from lung, colorectal and stomach cancer. We addressed this question in the dataset of the Seven Countries Study [5], in which the variation in intake of antioxidant (pro)vitamins and in cancer mortality is large. Furthermore, α-carotene, ß-carotene, vitamin C, and α-tocopherol were chemically analyzed in food equivalent composites representing usual intake of each of the 16 cohorts of the Seven Countries Study, thus avoiding measurement error caused by less reliable food composition tables.

Material and methods

Subjects

Between 1958 and 1964 more than 12,000 men aged 40 to 59 years were enrolled in the Seven Countries Study. In these countries, 16 cohorts were established: 11 among rural areas in Finland, Italy, Greece, the former Yugoslavia and Japan; 2 cohorts of railroad employees, one in the north western part of the USA and one in Rome; one of workers in a large cooperative in Serbia; one of faculty members of the Belgrade University; and one of inhabitants of a small commercial market town in the Netherlands. The characteristics of the cohorts have been described in detail by Keys *et al.* [5].

Dietary intake assessment

Dietary information was collected in small random samples (8-49 men) of the individual cohorts, using the 7-day record method in 14 of the cohorts, a 4-day record

in the Japanese cohort at Ushibuka, and a 1-day record in the cohort of the US Railroad employees. In 14 cohorts, this took place between 1959 and 1964, whereas in 2 cohorts dietary information was gathered around 1970. Information on vitamin supplements was not collected, but their use was uncommon in the 1960s. In 1986 the original dietary data of all these cohorts were recoded by one dietician (A.J.) in a standardized way, and the average daily food intake was calculated for each cohort [6].

Subsequently, foods representing the average food intake of each cohort at baseline were bought locally and sent by air in cooling boxes to the laboratory of the Department of Human Nutrition, Wageningen Agricultural University, The Netherlands (M.B. Katan). Within one day after arrival, the foods were cleaned and combined into food equivalent composites representing the average food intake of each cohort. These composites were subsequently homogenized, freeze-dried, and stored at -20 °C until analyzed. To a part of the food equivalent composites that was not frozen oxalic acid was added to preserve vitamin C. The antioxidant (pro)vitamin contents were determined at the State Institute for Quality Control of Agricultural Products, Wageningen, The Netherlands (P.C.H. Hollman). Determination of α and ß-carotene was done using HPLC followed by spectrophotometric measurements [7], and α-tocopherol was determined using HPLC with spectrofluorescence detection [8]. Vitamin C was analyzed fluorometrically [9] within 10 days after arrival. With these data the average intake of nutrients was calculated per cohort.

Vital status

The vital status of all men was checked almost every 5 years during 25 years of follow-up, and the primary cause of death of the men who died was established centrally by H. B. and A. M. In total, 56 men were lost to follow-up (0.4% of the entire study population). The end points in the present study are mortality from lung cancer (ICD 162), colorectal cancer (ICD 152-154) and stomach cancer (ICD 151) (8th Revision of the International Classification of Diseases). Mortality as a proportion of the initial population was highly correlated to mortality rates ($r>0.95$). Mortality as a proportion is reported because it can be more easily interpreted than number of deaths per person-year. The 25-year mortality was age-standardized using the direct method with the age distribution of the whole study population as a standard.

Statistical analyses

Regression analysis (PROC REG, SAS statistics), with the age-adjusted 25-year cancer mortality as dependent variable and the population average antioxidant (pro)vitamin intake as independent variable were first carried out univariately. Thereafter, potential confounders were added to the models as independent variables. The potential confounders consisted of known risk factors for the 3 types of cancer that were available in the dataset and of the other antioxidant (pro)vitamins. For lung cancer mortality the percentage of cigarette smokers and the intake of saturated fatty acids were considered as confounders [10]. Intake of dietary fibre and fat were considered as potential confounders for colorectal cancer mortality [11] and sodium or nitrate intake and percentage of cigarette smokers for stomach cancer mortality [12,13]. This was followed by adding the other antioxidant (pro)vitamins to the model. Due to the limited number of degrees of freedom in the statistical analyses (n=16), only one other antioxidant (pro)vitamin was adjusted for at the time, resulting in regression models with a maximum of 4 independent variables. Two-sided *p*-values less than 0.05 were considered statistically significant. Log-transformed variables were also used for all the above analyses, but since the results were similar untransformed results are presented.

Results

Mean intake of the antioxidant (pro)vitamins varied considerably by cohort (table 1). For α-carotene the range between the lowest and the highest intake was more than 30-fold, whereas for the other 3 antioxidants this range was 7- to 8-fold. The Serbian cohort in Velika Krsna and the Italian cohort in Crevalcore had a low average intake of α- and ß-carotene and vitamin C, whereas the US railroad cohort had a high intake of these antioxidants. The average intake of α-tocopherol was low in both Japanese cohorts and high in both Greek cohorts. Pearson correlation coefficients between the population average antioxidant (pro)vitamins ranged from -0.45 between α-carotene and α-tocopherol up to 0.39 between vitamin C and α-tocopherol. Dietary fibre was positively correlated with α-tocopherol (r=0.52), nitrate intake was positively related and sodium intake from foods was inversely related to ß-carotene intake (r=0.66, -0.53 respectively). The other potential confounders, intake of fat, saturated fatty acids, and percentage of cigarette smokers were not strongly correlated with any of the antioxidants studied.

Table 1. *Daily average antioxidant (pro)vitamin intake (mg) in the 16 cohorts of the Seven Countries Study.*

Country	Cohort	α-carotene	ß-carotene	vitamin C	α-tocopherol
Croatia	Dalmatia	0.03	4.21	60	15.9
Croatia	Slavonia	0.13	3.41	41	11.2
Serbia	Velika Krsna	0.03	0.62	17	7.8
Serbia	Zrenjanin	0.42	2.15	112	12.1
Serbia	Belgrade	0.62	1.74	70	18.3
Greece	Corfu	0.02	2.16	125	31.2
Greece	Crete	0.02	1.84	136	21.4
Italy	Rome railroad	0.02	2.64	53	14.3
Italy	Crevalcore	0.02	1.22	50	15.7
Italy	Montegiorgio	0.14	2.87	44	13.2
The Netherlands	Zutphen	0.30	2.90	110	8.6
Finland	West	0.71	2.14	65	9.1
Finland	East	0.22	1.43	80	9.6
USA	Railroad	0.46	2.57	142	6.8
Japan	Tanushimaru	0.43	1.89	39	4.7
Japan	Ushibuka	0.41	1.41	45	6.3

During the 25-year follow-up period, 1580 men died of cancer: 424 of lung cancer, 267 of stomach cancer, and 130 of colorectal cancer. The age-adjusted 25-year lung cancer mortality was lowest (1%) in Montegiorgio and Tanushimaru and highest in East Finland (7.3%) and Zutphen (7.2%) (figure 1). The highest 25-year stomach cancer mortality was observed in both Japanese cohorts (5.1%), whereas in the Belgrade cohort it was only 0.2%. The highest colorectal cancer mortality, 2%, was observed in Zutphen and among the US-railroad employees, and the lowest of 0.1% in East Finland.

None of the studied antioxidant (pro)vitamins was significantly related to 25-year lung cancer mortality (table 2) in univariate regression analyses. Adjustment for the percentage of smokers and intake of saturated fatty acids did not change the results, nor did further adjustment for the other antioxidant (pro)vitamins.

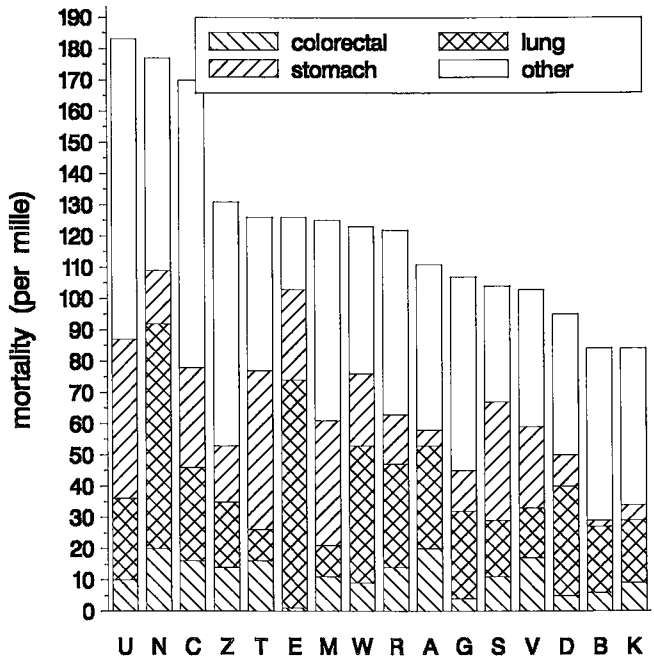


Figure 1. *25-year cancer mortality (‰) in 16 cohorts of the Seven Countries Study. A=US railroad; B=Belgrade; C=Crevalcore; D=Dalmatia; E=East Finland; G=Corfu; K=Crete; M=Montegiorgio; N=Zutphen; R=Rome railroad; S=Slavonia; T=Tanushimaru; U=Ushibuka; V=Velika Krsna; W=West Finland; Z=Zrenjanin*

Table 2. *Relations between average daily intake of antioxidant (pro)vitamins in 1960 and 25-year age-adjusted mortality (%) from lung cancer in the 16 cohorts of the Seven Countries Study*

Antioxidant	Univariate				Multiple[a]			
	intercept	beta[b]	P-value	R^2	intercept	beta[b]	P-value	R^2
α-carotene (0.01 mg)	2.78	0.01	0.60	0.02	5.13	-0.00	0.77	0.64
ß-carotene (0.1 mg)	2.70	0.02	0.77	0.01	4.82	0.01	0.79	0.64
vitamin C (10 mg)	1.98	0.15	0.26	0.09	4.19	0.11	0.21	0.69
α-tocopherol (mg)	3.56	-0.04	0.61	0.02	4.66	0.07	0.20	0.69

[a] Multiple regression adjusted for percentage smokers and saturated fat intake

[b] Regression coefficient

Table 3. *Relations between average daily intake of antioxidant (pro)vitamins in 1960 and 25-year age-adjusted mortality (%) from stomach cancer in the 16 cohorts of the Seven Countries Study*

Antioxidant	Univariate				Multiple[a]			
	intercept	beta[b]	P-value	R^2	intercept	beta[b]	P-value	R^2
α-carotene (0.01 mg)	2.31	0.00	0.93	0.00	5.13	0.00	0.89	0.30
ß-carotene (0.1 mg)	3.14	-0.04	0.45	0.04	6.67	-0.04	0.46	0.33
vitamin C (10 mg)	4.33	-0.27	0.01	0.44	7.78	-0.28	0.0003	0.77
α-tocopherol (mg)	3.88	-0.12	0.04	0.26	5.30	-0.10	0.09	0.45

[a] Multiple regression adjusted for sodium intake and percentage smokers
[b] Regression coefficient

Table 4. *Relations between average daily intake of antioxidant (pro)vitamins in 1960 and 25-year age-adjusted mortality (%) from colorectal cancer in the 16 cohorts of the Seven Countries Study*

Antioxidant	Univariate				Multiple[a]			
	intercept	beta[b]	P-value	R^2	intercept	beta[b]	P-value	R^2
α-carotene (0.01 mg)	1.12	0.00	0.90	0.00	2.31	-0.00	0.66	0.37
ß-carotene (0.1 mg)	1.24	-0.00	0.80	0.00	2.43	-0.01	0.47	0.38
vitamin C (10 mg)	1.16	-0.00	0.96	0.00	2.13	0.02	0.50	0.38
α-tocopherol (mg)	1.70	-0.04	0.04	0.26	2.23	-0.02	0.32	0.41

[a] Multiple regression adjusted for intake of dietary fibre and fat
[b] Regression coefficient

Average vitamin C intake was strongly and inversely related to the stomach cancer mortality in univariate regression analysis (figure 2), and explained 44% of its variance. After adjustment for sodium intake and percentage of smokers, a 10 mg higher intake of vitamin C was associated with 0.28% lower 25-year stomach cancer mortality (table 3). This model explained 77% of the variation in stomach cancer mortality, and only percentage of smokers and the vitamin C intake were significant contributors to the model. Replacing sodium intake by nitrate intake did not change these results. Further adjustment for the other antioxidant (pro)vitamins did not alter the results, and the regression coefficient and its statistical significance were not affected by the removal of any single cohort. In univariate analysis, α-tocopherol was inversely related to the stomach cancer mortality and explained 26% of its variance. This relation weakened somewhat and became borderline significant after adjustment

for the percentage of smokers and sodium intake. After further adjustment for vitamin C intake, this relation disappeared (ß=-0.03% per mg α-tocopherol; p=0.52), as vitamin C intake was positively related to α-tocopherol intake, and inversely to stomach cancer mortality.

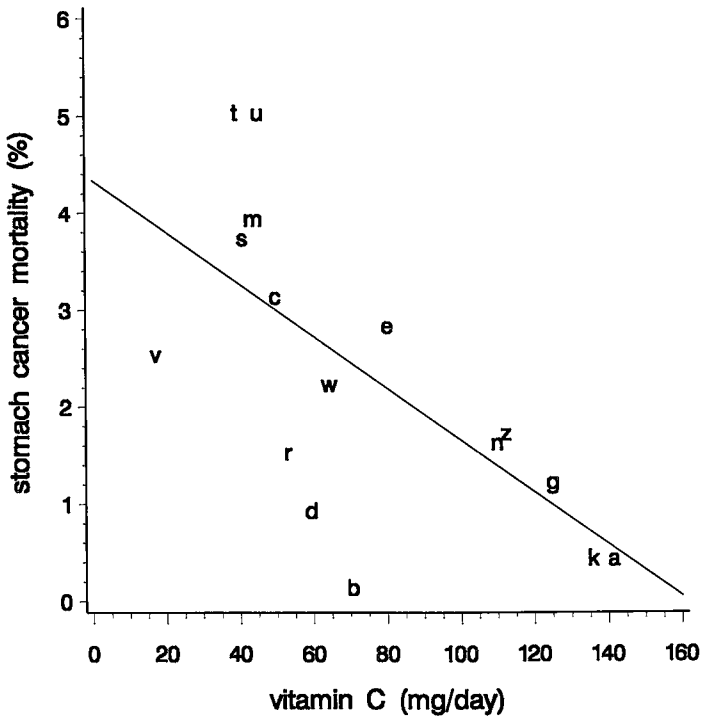


Figure 2. *Univariate association between vitamin C intake at baseline and 25-year age-adjusted mortality from stomach cancer in the Seven Countries Study.*
A=US railroad; B=Belgrade; C=Crevalcore; D=Dalmatia; E=East Finland; G=Corfu; K=Crete; M=Montegiorgio; N=Zutphen; R=Rome railroad; S=Slavonia; T=Tanushimaru; U=Ushibuka; V=Velika Krsna; W=West Finland; Z=Zrenjanin

In univariate regression analyses, average intake of α-tocopherol was inversely related to 25-year colorectal cancer mortality and explained 26% of its variance (table 4). However, after adjustment for intake of dietary fibre and fat this relation almost completely disappeared, due to the confounding effect of dietary fibre which was positively related to α-tocopherol intake and inversely to colorectal cancer mortality. Neither carotenoids nor vitamin C were related to colorectal cancer mortality. Adjustment for other antioxidants did not alter the relationships.

Discussion

In the Seven Countries Study, average population intake of vitamin C was strongly inversely related to the 25-year stomach cancer mortality of the 16 cohorts, but not to mortality from lung and colorectal cancer. The average intake of α- and ß-carotene, and α-tocopherol were not independently related to mortality from lung, stomach and colorectal cancer.

In interpreting these results, it should be considered that this study is by design an ecological study. The advantages of this design were already mentioned: a large variation in both exposure and outcome, and relatively small measurement errors in the exposure. The design, however, has several drawbacks. For dietary intake, only information on population averages rather than whole distributions, is available, which means that only linear, no-threshold associations can be studied and that adjustment for confounders or interaction effects can only be done crudely. It is also likely that for some factors important in the etiology of the disease studied, that differ considerably across populations, no information is available. Because of these drawbacks, it is often concluded that causal inferences cannot be drawn from ecological studies [14]. Furthermore, true relationships on the population level are not necessarily the same as those on the individual level within populations. For example, when a population is homogeneous with respect to a certain determinant of a disease, while this determinant varies considerably between populations, the relationship can be found only at population level [15]. It is therefore important that results from ecological studies be judged in the light of plausible biological mechanisms rather than on consistency with the results of cohort and case-control studies.

Two more disadvantages of the Seven Countries Study should receive attention, although we expect that their influence on the results is small. Firstly, there was a 25-year period between the collection of dietary information and the buying of the foods

to be analyzed. Changes in food composition during the period could lead to bias. Secondly, differences between countries in diagnosis and treatment of the 3 types of cancer could result in relationships to mortality data that are not valid for incidence.

Our results show that the average population intake of vitamin C was inversely related to stomach cancer mortality, and that this relation was independent of the other potential risk factors (percentage of smokers, and average intakes of sodium and nitrate). In relation to sodium intake, residual confounding may play a role, since sodium intake did not include salt added during preparation or at table. We could not determine whether the inverse relation was also independent from infections with *Helicobacter pylori*, as no information on this infection is available for the Seven Countries Study. It seems, however, that infection with *H. pylori* cannot be a strong confounder, as its prevalence is high in Japan and Greece and low in the USA and Italy [16], whereas the average vitamin C intake is low in the Seven Countries cohorts in Japan and Italy, and high in those in the USA and Greece.

In other international ecological studies on stomach cancer only associations with food group intake were reported. In contrast with our study, fruit intake as an indicator for vitamin C intake was not related to stomach cancer mortality in 4 studies [17-20], and weakly positively associated in one study [21]. For vegetable intake, another source of vitamin C, no association [17-21] is reported in other correlation studies. A possible explanation for this discrepancy could be that the dietary intake data of the correlation studies referred to are *per capita* disappearance data (food balance sheets), which are only indirectly related to dietary intake and are likely to be of variable quality [22]. A second disadvantage of these studies is the fact that exposure and outcome do not apply to the same population, whereas in the Seven Countries Study, prospectively collected data were used in the ecological analyses.

We cannot exclude the possibility that another factor or a combination of other factors highly associated with vitamin C intake is etiologically related to stomach cancer mortality. However, an inverse association between intake of vitamin C and stomach cancer mortality fits the Correa model for intestinal-type stomach cancer [23]. This model hypothesizes that stomach cancer develops through a sequence of histological changes: the normal mucosa is affected by superficial gastritis, leading to chronic atrophic gastritis, intestinal metaplasia, epithelial dysplasia, finally resulting in carcinoma and invasion of surrounding tissue. In the step from chronic atrophic gastritis to the later stages, higher pH, bacterial growth and reduction of nitrate to

nitrite are important. Nitrosating species derived from nitrite can react with nitrosable compounds such as amines and amides and form carcinogenic *N*-nitroso compounds. Ascorbic acid is able to inhibit the *in vitro* nitrosation of different classes of compounds, and thus appears to have potential importance as an *in vivo* nitrite scavenger [24]. Our finding, that nitrate intake was not a significant contributor in the multiple regression model, could be explained by the fact that nitrate exposure is not the rate limiting factor in the endogenous formation of *N*-nitroso compounds as was suggested by Forman [13].

In the present study, none of the 4 antioxidants studied was associated with lung cancer mortality. The percentage of smokers and the average saturated fatty acid intake, the 2 potential confounders, were each significant in the linear regression models and explained about 60% of the variance in lung cancer mortality. We are not aware of other international correlation studies reporting the relationship between antioxidant (pro)vitamins and lung cancer. An indirect comparison can be made, however, with a study by Hursting *et al.* [25], who found no association between lung cancer and intake of polyunsaturated fat (PUF). As PUF is an indicator for vitamin E intake, this is in line with our results. Three international studies reported no association between fruit intake and lung cancer mortality [17-19]. For vegetable intake a weak inverse association [18] and no association [17,19] were reported. Overall, this supports our results of no association of α-carotene, ß-carotene, and vitamin C with lung cancer mortality, on the ecological level.

In the Seven Countries Study, the average population intake of α-carotene, ß-carotene, vitamin C, or α-tocopherol could not independently explain colorectal cancer mortality. For α-tocopherol an inverse relation was observed in univariate regression analyses, but this disappeared after adjustment for dietary fibre with which it had a correlation of 0.52. Also, in the study by McKeown-Eyssen and Bright-See [11] the availability of vitamin C in 38 countries was not inversely associated with colon cancer death rate. Weak positive [17] and weak inverse association [18], and no association [21,26] were reported in international studies on vegetable intake and cancer of the colon and/or rectum. Also for fruit intake, weak positive [17], and weak inverse association [21], and no association [18,26] were found. Four international studies on colon and/or rectal cancer mortality observed no association with either vegetable fat [11,26] or PUF [25,27], both associated with vitamin E. Overall, on the

ecological level, the variability in colorectal cancer mortality cannot be explained by differences in intake of antioxidant (pro)vitamins.

In conclusion, the inverse relation between vitamin C and stomach cancer previously observed in case-control and cohort studies was confirmed on the population level by the comparisons of cohorts in the Seven Countries Study. Despite the large variations in cancer mortality and in antioxidant intake, there was no indication of a protective effect of carotenoids and α-tocopherol on cancer of the lung, stomach and colorectum, or of vitamin C on lung and colorectal cancer.

Acknowledgements

The mortality follow-up of the 2 cohorts in Dalmatia and Slavonia was supported by a grant from the Netherlands Nutrition Foundation to D. K. We thank Dr. A. Keys for his drive and assiduity in carrying the study through for more than 25-years. We are grateful to Dr B.P.M. Bloemberg, Dr S. Giampaoli, H.E. Goddijn, Dr R. Schlemper, and Dr W.M.M. Verschuren, for their contribution to the collection and preparation of the food equivalent composites, to P. van den Bovenkamp and Dr. M.B. Katan, Department of Human Nutrition, Wageningen Agricultural University, and P.C.H. Hollman, State Institute for Quality Control of Agricultural Products Wageningen for preparation and vitamin analyses of the food equivalent composites. Dr. H.B. Bueno de Mesquita, and Dr. W.A. van Staveren are thanked for their helpful comments on an earlier version of the manuscript.

References

1. Tomatis L, Aitio A, Day NE, Heseltine E, Kaldor J, Miller AB, Parkin DM, Riboli E (eds). Cancer: causes, occurrence and control. IARC Scientific publications, 100. Lyon: International Agency for Research on Cancer, 1990
2. Block G. Vitamin C and cancer prevention: the epidemiologic evidence. Am J Clin Nutr 1991; 53: 270s-282s
3. Van Poppel GAFC. Carotenoids and cancer: an update with emphasis on human intervention studies. Eur J Cancer 1993; 29A: 1335-1344
4. Knekt P, Aromaa A, Maatela J, Aaran R-K, Nikkari T, Hakama M, Hakulinen T, Peto R, Teppo L. Vitamin E and cancer prevention. Am J Clin Nutr 1991; 53: 283s-286s
5. Keys A and 13 others. Epidemiological studies related to coronary heart disease: characteristics of men aged 40-59 in seven countries. Acta Med Scand 1967; 460(suppl): 1-392
6. Kromhout D, Keys A, Aravanis C, Buzina R, Fidanza F, Giampaoli S, Jansen AM, Menotti A, Nedeljkovic S, Pekkarinen M, Simic BS and Toshima H. Food consumption patterns in the 1960s in seven countries. Am J Clin Nutr 1989; 49: 889-894

7. Speek AJ, Temalilwa CR, Schrijver J. Determination of ß-carotene content and vitamin A activity of vegetables by high-performance liquid chromatography and spectrophotometry. Food Chemistry 1986; 19: 65-74
8. McMurray CH, Blanchflower WJ. Determination of α-tocopherol in animal feedstuffs using high-performance liquid chromatography with spectrofluorescence detection. J Chromatogr 1979; 176: 488-492
9. Roy RB, Conetta A, Salpeter J. Automated fluorometric method for the determination of total vitamin C in food products. J Assoc Off Anal Chem 1976; 59: 1244-1250
10. Xie J, Lesaffre E, Kesteloot H. The relationship between animal fat intake, cigarette smoking, and lung cancer. Cancer Causes Control 1991; 2: 79-83
11. McKeown-Eyssen GE, Bright-See E. Dietary factors in colon cancer: international relationships. Nutr Cancer 1984; 6: 160-170
12. Howson CP, Hiyama T, Wynder EL. The decline in gastric cancer: epidemiology of an unplanned triumph. Epidemiol Rev 1986; 8: 1-27
13. Forman D. The etiology of gastric cancer. In: O'Neill IK, Chen J, Bartsch H (eds). Relevance to human cancer of N-nitroso compounds, tobacco smoke and mycotoxins. Lyon: IARC, 1991, pp 22-32
14. Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health 1982; 72: 1336-1344
15. Rose G. Sick individuals and sick populations. Int J Epidemiol 1985; 14: 32-38
16. Eurogast Study Group. An international association between Helicobacter pylori infection and gastric cancer. Lancet 1993; 341: 1359-1362
17. Howell MA. Factor analysis of international cancer mortality data and per capita food consumption. Br J Cancer 1974; 29: 328-336
18. Schrauzer GN. Cancer mortality correlation studies. II. Regional associations of mortalities with the consumptions of foods and other commodities. Med Hypotheses 1976; 2: 39-49
19. Maruchi N, Aoki S, Tsuda K, Tanaka Y, Toyokawa H. Relation of food consumption to cancer mortality in Japan, with special reference to international figures. Gann 1977; 68: 1-13
20. Tominaga S, Ogawa H, Kuroishi T. Usefulness of correlation analyses in the epidemiology of stomach cancer. Natl Cancer Inst Monogr 1982; 62: 135-140
21. Knox EG. Foods and diseases. Br J Prev Soc Med 1977; 31: 71-80
22. Willett WC. Epidemiologic studies of diet and cancer. Med Oncol Tumor Pharmacother 1990; 7: 93-97
23. Correa P. A human model of gastric carcinogenesis. Cancer Res 1988; 48: 3554-3560
24. Tannenbaum SR, Wishnok JS, Leaf CD. Inhibition of nitrosamine formation by ascorbic acid. Am J Clin Nutr 1991; 53: 247s-250s
25. Hursting SD, Thornquist M, Henderson MM. Types of dietary fat and the incidence of cancer at five sites. Prev Med 1990; 19: 242-253
26. Rose DP, Boyar AP, Wynder EL. International comparisons of mortality rates for cancer of the breast, ovary, prostate, and colon, and per capita food consumption. Cancer 1986; 58: 2363-2371
27. Liu K, Stamler J, Moss D, Garside D, Persky V, Soltero I. Dietary cholesterol, fat, and fibre, and colon-cancer mortality. An analysis of international data. Lancet 1979; ii: 782-785

Chapter 3

Repeated measurements of vegetables, fruits, and antioxidant (pro)vitamins in relation to lung cancer (The Zutphen Study)

Marga Ocké, Bas Bueno de Mesquita, Edith Feskens,
Wija van Staveren, and Daan Kromhout

submitted for publication

Abstract

The authors studied the intake of vegetables, fruits, ß-carotene, vitamin C and E in relation to incidence of lung cancer. For 561 men from the town of Zutphen, the Netherlands, dietary history information was obtained in 1960, 1965, and 1970. During 1971-1990 54 new cases of lung cancer were identified. The data were analyzed using Cox proportional hazard analyses, adjusting for age, pack-years of cigarettes, and energy intake. No relationship between intake of vitamin E and lung cancer risk was seen. For vitamin C intake the results pointed to an inverse association, although not entirely consistently. Furthermore, it was observed that subjects with low stable intakes (i.e. low in 1960, 1965, and 1970) of vegetables, fruits, and ß-carotene experienced more than two-fold increased relative risks on lung cancer than those with high stable intakes. For subjects with low average intakes relative risks were much lower and not statistically significant. In conclusion, vitamin E seems not related to lung cancer risk, whereas for ß-carotene, vitamin C, vegetables, and fruit most studies, including the present one, suggest weak inverse associations. The use of repeated intake measurements to select subgroups with stable, highly contrasting intakes may be a promising approach for studying diet-cancer relationships.

Introduction

Worldwide, lung cancer is the leading cause of cancer mortality among both men and women. Also among men in the Netherlands, the lung cancer incidence and mortality rates are high, i.e. 100.2 respectively 95.9 per 100,000 person-years in 1991 [1]. Cigarette smoking is the most important cause of lung cancer, but other factors like diet may also be involved [2].

A large proportion of the observational studies have confirmed an inverse relationship between the risk of lung cancer and the intake of vegetables and fruits [3]. However, associations found in cohort studies are less strong than in case-control studies [4], which raises the possibility of recall or publication bias in case-control studies. From the thousands of compounds available in vegetables and fruits, a large number is potentially anticarcinogenic. Since Peto *et al.* [5] postulated that a high intake of ß-carotene might decrease the risk of cancer in humans, considerable interest has focused on antioxidant (pro)vitamins and cancer prevention. Previous prospective cohort studies on intake of these antioxidants and lung cancer suggest a protective effect of ß-carotene, although often not statistically significant, whereas results are inconclusive for the vitamins C and E [6]. Also, in a meta-analysis of serum and plasma studies, Comstock *et al.* [7] conclude that ß-carotene levels are consistently inversely related to lung cancer risk, whereas for vitamin E results are less clear. In one study that related plasma vitamin C levels to lung cancer mortality no association was observed [8].

Two intervention studies on the relation between antioxidant (pro)vitamins and lung cancer didn't shed more light on this topic either. In the first in Linxian, China, the risk of death from lung cancer was reduced by 45 percent ($p = 0.11$) among those receiving the supplement with ß-carotene, α-tocopherol, and selenium [9]. In a second trial, among Finnish male smokers 50 to 69 years of age, no reduction in lung cancer incidence was observed among the men who received α-tocopherol, whereas a higher incidence among the men who received ß-carotene was found [10]. The results of the ATBC-trial should not, however, be seen as proving these antioxidants to be ineffective or even hazardous, since the trial was of limited duration in relation to the lifetime exposure to cigarette smoke and other carcinogens and the daily dose of vitamin E was rather small [11]. So, at present the question whether antioxidant (pro)vitamins reduce the incidence of lung cancer remains unanswered.

Antioxidant (pro)vitamins probably retard some types of initiation and quite clearly inhibit the promotion process [12]. Continuous and abundant presence of these antioxidants seems therefore highly important. In most studies, however, estimates of intake are based on a single dietary assessment, usually a dietary history interview or a food frequency questionnaire. Although these types of dietary assessment methods try to assess the habitual diet, the period of reference is generally short, ranging from a few months to a year. In addition, substantial measurement error is made in assessing recent diet. The assumption that a single measurement represents the long-term habitual diet may therefore be erroneous, particularly in case of changes in the dietary pattern [13]. In fact, large changes in the diet have been observed in most western populations during the last decades. Likewise, the men in the prospective Zutphen Study substantially changed their diets after the start of the study in 1960 [14]. For this reason, we studied the relationship between the intakes of vegetables, fruits, ß-carotene, vitamin C, and vitamin E and the incidence of lung cancer in the Zutphen Study taking into account intake data obtained at three points over time.

Materials and methods

Subjects

The design of the Zutphen Study has been described elsewhere [15]. Briefly, it is a prospective cohort study on risk factors for chronic diseases and forms the Dutch contribution to the Seven Countries Study. In 1960, 1088 men born between 1900 and 1919 and residing for at least five years in the town of Zutphen, the Netherlands, were invited for the study. Of these 1088 men, 872 participated in the baseline dietary survey, filled in a questionnaire and underwent a physical examination. The data were collected repeatedly, among others in 1965 and 1970. Some subjects did not respond to the invitation to participate in these latter surveys (n = 241) and 41 subjects died before 1970. Consequently, dietary information for 1960, 1965 and 1970 was available for 590 men. From these, 22 subjects were excluded because information on smoking in 1970 was missing, and 14 subjects were excluded who had a history of cancer in 1970.

Dietary data

In 1960, 1965 and 1970 cross-check dietary history interviews adapted to the Dutch situation were conducted by extensively trained dieticians. The interviews took place

during the spring and early summer, and concerned the habitual food consumption in the preceding 6-12 months. The wives of the participants were interviewed for about one hour about the participants' food consumption pattern at home and the participants were interviewed for about 20 minutes about their food consumption pattern outside the home. At that time research interest in vegetables and fruits was considerably less, which was reflected in the level of detail of the questions. For vegetables (excluding potatoes) the consumption frequency was asked of boiled and raw vegetables and traditional Dutch mixed dishes of vegetables, potatoes and meat. For fruits the consumption frequency was asked for citrus fruits, other fruits, apple compote and cooking pear. The habitual portion sizes were estimated by the dieticians according to the description of the interviewee and those difficult to estimate were measured on a post office scale. The cross-checks consisted of the comparison with the average of foods consumed during a day or a week and with the quantities of foods bought per week for the whole family. The dieticians also recorded the use of vitamin supplements. The data were recoded in 1978/1979 by two dietitians using the Dutch uniform food encoding system [16]. The food intake data were converted into energy and nutrients using the Dutch food table containing the composition of foods in the 1960s [17]. The ß-carotene and vitamin E content of foods were derived from more recent sources [18].

Other data

Information on among others occupation and smoking habits was collected by means of a standardized questionnaire. The number of packs of cigarettes smoked per day and the number of years of smoking were multiplied to obtain the variable pack-years of cigarettes. The information on occupation was grouped into four categories of socio-economic status.

Follow-up

In 1985 and 1990, the vital status of all men was obtained from the municipal population registries. No man was lost to follow-up. Information about the cause of death was obtained from the Central Bureau for Statistics and verified by means of hospital discharge and cancer registry data, and information from the general practitioner. Prevalence of a history of cancer was recorded during the medical examinations in 1985 and 1990 with a standardized questionnaire, and verified with

written information from the subject's general practitioner, with hospital discharge data, and information of the cancer registry. All morbidity data were checked and uniformly coded using strict criteria by two physicians and a nurse. The year of first clinical diagnosis of the various chronic diseases was recorded. In the present analysis, data on the occurrence of lung cancer and total mortality during 1971-1990 were included.

Data analysis

The computations were performed using the statistical package SAS, release 6.10. All statistical testing was two-sided; p values < 0.05 were considered to be statistically significant. The Student t and chi-square tests were used to test whether cases and non-cases differed in characteristics as measured in 1970. Repeated measures analysis of variance was performed with the three measurements of either vegetable, fruit, or an antioxidant (pro)vitamin intake as dependent variable, and time, disease status, and their interaction as independent variables. When the time-effect was statistically significant a test for linear trend was done. Pearson correlation coefficients between intake measurements of interest were calculated.

Cox proportional hazards models were used to estimate the relative risks of different levels of vegetable, fruit, and antioxidant (pro)vitamin intake to 20-year lung cancer incidence. This was first done with intake data from the separate years, with three levels of intake. For each year the same cut-off levels were chosen: the 33th and 67th percentile of the average intake over 1960, 1965 and 1970. Subjects who had used multivitamin supplements were placed in the highest intake level of vitamins E and C; those who had used vitamin C supplements only were also placed in the latter category. Thereafter, categorization of the study population was done for two types of combinations of the repeated measurements. In the first approach, the averages over the three repeated measures were calculated. Two categories were then formed: those with an average intake below and those with an average intake above the 33th percentile. In the second approach again two categories were created, but now from the subjects whose intake was either below or above the 33th percentile in all three years. Together, the subjects with a low or high stable intake consisted of about half of the study population, which was also the reason why only two categories were formed. In all models, the category with highest intake served as reference category. The continuous variables age and pack-years of cigarettes in 1970 and energy intake in the same year as the other intake measurements were included as potential confounders.

The adequacy of the proportional hazard assumption was checked by visual inspection of survival curves and testing time-dependent covariates [19]. Chi-square values for trend were calculated to determine whether dose-response relations were present.

Results

During 20 years of follow-up (1971-1990) 54 new cases of lung cancer were diagnosed and 269 of the 561 men died. The incidence density of lung cancer was 634 per 100,000 person-years; the median interval to diagnosis was 12.5 years of follow-up. In table 1, the incident lung cancer cases and non-cases are compared with respect to characteristics in 1970. The number of subjects using vitamin supplements was small. The majority of the study population consisted of current smokers (52 percent) and ex-smokers (40 percent). A larger proportion of the cases smoked and on average they had also smoked more cigarettes during their lifetime. The relative risk of subjects in the highest tertile of packyears of cigarettes was 4.37 (95 percent confidence interval 1.90-10.04) compared to those in the lowest tertile.

Table 1. *Selected characteristics in 1970 of 561 middle-aged men according to the incidence of lung cancer during 1971-1990: The Zutphen study.*

	Non-cases		Cases	
	mean	sd[a]	mean	sd
Number	507		54	
Age (years)	59.5	5.4	59.3	5.1
Current smokers (%)	49		74[b]	
Packyears of cigarettes	21.5	16.7	27.4[b]	143
Socio-economic status (%)				
high professionals	20		7	
small self-employed	19		22	
low, white collar	26		30	
low, blue collar	35		41	
Supplement usage (%)				
vitamin C	6		6	
multivitamins	5		0	

[a] standard deviation

[b] $p \leq 0.05$ for Students t-test or chi-square test

Average intakes of selected dietary components in 1960, 1965 and 1970 are presented in table 2. Total energy intake decreased over time which was mainly due to decreases in the intakes of potatoes, bread and added fats (data not shown). Average intakes of vitamin E, ß-carotene and vegetables also decreased over time, whereas the intake of fruits increased. Intake of vitamin C remained about the same, although the source of vitamin C shifted from vegetables and potatoes towards fruits. Incident cases had a lower average intake of vitamin C than non-cases which was most pronounced in 1970. For none of the dietary components did cases and non-cases differ in time-effect.

Table 2. *Average daily intake of energy, antioxidant (pro)vitamins, vegetables, and fruits in 1960, 1965 and 1970 for 561 middle-aged men according to the incidence of lung cancer during 1971-1990: The Zutphen Study*

		1960		1965		1970	
		mean	sd[a]	mean	sd	mean	sd
energy (MJ)[b]	Non-cases	13.0	2.7	12.4	2.7	10.8	2.2
	Cases	13.0	2.2	12.9	3.1	11.6	2.0
ß-carotene (mg)[b]	Non-cases	1.30	0.45	1.14	0.38	1.18	0.32
	Cases	1.23	0.51	1.14	0.38	1.16	0.39
vitamin C (mg)[c]	Non-cases	93	33	92	40	97	46
	Cases	86	36	89	31	79	34
vitamin E (mg)[b]	Non-cases	19.9	5.8	18.2	6.81	5.4	6.4
	Cases	19.3	5.2	18.4	6.71	6.1	5.8
vegetables (g)[b]	Non-cases	203	73	177	69	182	58
	Cases	192	78	176	74	171	64
fruits (g)[b]	Non-cases	117	85	153	111	171	130
	Cases	107	90	146	92	136	128

[a] standard deviation; [b] $p < 0.05$ for linear trend over time ; [c] $p < 0.05$ for lung cancer status effect

The correlation coefficients between the repeated measurements of the antioxidant (pro)vitamins, vegetables and fruits ranged between 0.25 and 0.44 (table 3). Intake of vitamin C nor ß-carotene correlated with vitamin E in any of the years (r =-0.06 to 0.11). Beta-carotene intake was strongly correlated with intake of vegetables (r = 0.88 - 0.91), and intake of vitamin C with intake of fruits (r = 0.68 - 0.81). The correlation between the intakes of ß-carotene and fruits decreased over time from 0.30 to 0.07, and that between vitamin C and vegetables from 0.58 to 0.23. As

a result the correlation between vitamin C and ß-carotene intake also decreased over time: from 0.60 in 1960 via 0.47 in 1965 to 0.23 in 1970. Correlation coefficients between intakes of vegetables and fruits were low in all years (0.01-0.21). In general, men with a high intake of vegetables, fruits, ß-carotene, or vitamin C, but not of vitamin E, had smoked less cigarettes during their life than those with a low intake (data not shown).

Table 3. *Pearson correlation coefficients between intakes of antioxidant (pro)vitamins, vegetables, and fruits in 1960, 1965, and 1970 for 561 middle-aged men: The Zutphen Study*

	1960/1965	1965/1970	1960/1970
ß-carotene	0.36	0.35	0.33
vitamin C	0.38	0.43	0.31
vitamin E	0.44	0.25	0.25
vegetables	0.29	0.30	0.33
fruits	0.37	0.44	0.29

Relative risks for levels of intake of antioxidant (pro)vitamins, vegetables and fruits in 1960, 1965, and 1970 adjusted for age, pack-years of cigarettes and energy intake are presented in table 4. About two-fold increased risks were observed for subjects in the lowest intake categories of vitamin C and fruits in 1970. There was also a significant dose-response relationship, although the relative risk for the middle category of fruit intake was below one. For vitamin C intake in 1960 the relative risk in the lowest level was 1.64, whereas that for 1965 did not exceed one. For fruits the relative risks in 1960 and 1965 were also lower than in 1970. The intakes of ß-carotene, vitamin E, and vegetables were not related to lung cancer risk in any of the single years. When not placing subjects that used multivitamin supplements in the category of high vitamin E intake, the relative risks for vitamin E were lower (1.18 for lowest category in 1970). Additional adjustment for socio-economic status or alcohol consumption did not materially alter the results, nor did adjustment for fat instead of energy intake in the models with vitamin E. Similarly, when the intakes of the three antioxidant (pro)vitamins or of vegetables and fruits were analyzed in one model, the results remained essentially the same. In none of the models, interaction terms between intake levels and packyears of cigarettes were statistically significant (all $p > 0.15$).

Table 4. *Adjusted[a] relative risks (95% confidence intervals) for antioxidant (pro)vitamin, fruit en vegetable intake in 1960, 1965, and 1970 in relation to incident lung cancer during 1971-1990: The Zutphen Study.*

		1960		1965		1970
ß-carotene (mg)						
< 1.07	1.35	(0.74-2.48)	0.82	(0.44-1.53)	1.40	(0.73-2.66)
1.07 - 1.31	0.81	(0.39-1.71)	0.99	(0.48-2.05)	0.93	(0.46-1.89)
> 1.31	1.00		1.00		1.00	
p_{trend}[b]		*0.33*		*0.50*		*0.28*
Vitamin C (mg)						
< 80	1.64	(0.88-3.07)	0.95	(0.50-1.82)	2.16	(1.14-4.09)
80 - 102	1.27	(0.61-2.61)	1.28	(0.66-2.49)	1.76	(0.80-3.87)
> 102	1.00		1.00		1.00	
p_{trend}		*0.12*		*0.88*		*0.02*
Vitamin E (mg)						
< 15.3	1.39	(0.62-3.13)	1.12	(0.55-2.28)	1.47	(0.66-3.17)
15.3 - 19.8	1.49	(0.74-2.99)	0.90	(0.44-1.82)	1.45	(0.66-3.17)
> 19.8	1.00		1.00		1.00	
p_{trend}		*0.39*		*0.77*		*0.34*
Vegetables (g)						
< 163	1.31	(0.71-2.41)	1.06	(0.57-2.00)	1.07	(0.56-2.03)
163 - 205	0.77	(0.38-1.57)	0.83	(0.38-1.81)	0.81	(0.40-1.62)
> 205	1.00		1.00		1.00	
p_{trend}		*0.43*		*0.77*		*0.79*
Fruits (g)						
< 107	1.33	(0.63-2.80)	1.20	(0.64-2.27)	1.92	(1.04-3.55)
107 - 166	1.15	(0.50-2.63)	0.92	(0.46-1.85)	0.91	(0.39-2.11)
> 166	1.00		1.00		1.00	
p_{trend}		*0.43*		*0.56*		*0.03*

[a] Obtained by Cox proportional hazard analyses, adjusted for age, packyears of cigarettes, and energy intake; [b] *p*-value for Chi-square test for trend

In table 5, subjects with low average and those with low stable intakes of vegetables, fruits, and antioxidant (pro)vitamins are compared on lung cancer risk, smoking, and intake with subjects with high average and high stable intakes, respectively. Relative risks regarding low stable intakes were more than two-fold for ß-carotene, vegetables and fruits, 1.65 for vitamin C and 1.54 for vitamin E. The relative risks for low average intakes were below 1.5 for all comparisons, and ranged from 1.04 for ß-carotene to 1.43 for vitamin C. The differences in average intakes of

the three antioxidant (pro)vitamins, vegetables, and fruits were 50 percent higher when subjects with low and high stable intakes were compared than in the comparisons of subjects with low and high average intakes. With the exception of the vitamin E subgroups, the larger contrasts in intake among subjects with stable intakes were accompanied by larger opposite contrasts in packyears of cigarettes smoked.

Table 5. *Adjusted[a] relative risks on incident lung cancer and characteristics of subjects with a low and those with a high average and stable intake of antioxidant (pro)vitamins, vegetables, and fruits*

	Average intake		Stable intake	
percentile:	≤33th	>33th	≤33th	>33th
ß-carotene				
N (# cases)[b]	188 (19)	373 (35)	72 (13)	178 (17)
Packyears of cig.[c]	23.2 ± 16.4[d]	21.6 ± 16.7	25.0 ± 16.9	21.4 ± 16.0
ß-carotene, mg	0.90 ± 0.15	1.36 ± 0.21	0.78 ± 0.16	1.47 ± 0.19
RR (95% ci)[e]	1.04 (0.60-1.83)	1.00	2.11 (1.02-4.38)	1.00
Vitamin C				
N (# cases)	148 (19)	413 (35)	81 (11)	211 (18)
Packyears of cig.	25.6 ± 16.7	20.9 ± 16.4	26.8 ± 16.7	19.7 ± 15.6
vitamin C, mg	63 ± 12	104 ± 27	55 ± 10	117 ± 29
RR (95% ci)	1.43 (0.82-2.51)	1.00	1.65 (0.76-3.58)	1.00
Vitamin E				
N (# cases)	166 (17)	395 (37)	55 (8)	206 (20)
Packyears of cig.	22.1 ± 17.1	22.1 ± 16.4	20.9 ± 17.2	20.4 ± 15.9
Vitamin E, mg	13.0 ± 1.9	19.9 ± 3.9	11.7 ± 2.02	1.8 ± 3.6
RR (95% ci)	1.37 (0.72-2.60)	1.00	1.54 (0.56-4.25)	1.00
Vegetables				
N (# cases)	186 (20)	375 (34)	65 (11)	183 (15)
Packyears of cig.	22.9 ± 16.4	21.7 ± 16.7	26.8 ± 17.7	20.3 ± 15.4
Vegetable, g	135 ± 24	212 ± 37	113 ± 26	229 ± 32
RR (95% ci)	1.19 (0.68-2.06)	1.00	2.13 (0.97-4.68)	1.00
Fruits				
N (# cases)	187 (22)	374 (32)	87 (14)	162 (12)
Packyears of cig.	25.9 ± 16.5	20.2 ± 16.3	26.2 ± 15.5	18.8 ± 15.0
Fruits, g	67 ± 30	184 ± 73	43 ± 25	221 ± 83
RR (95% ci)	1.39 (0.80-2.41)	1.00	2.52 (1.15-5.57)	1.00

[a] Obtained by Cox proportional hazard analyses, adjusted for age, packyears of cigarettes, energy intake; [b] Total number (number of cases); [c] Packyears of cigarettes; [d] mean ± standard deviation; [e] Relative risk (95% confidence interval)

Discussion

We studied the relationships between intakes of vegetables, fruits, and antioxidant (pro)vitamins and lung cancer incidence using exposure information obtained in different years and combinations of this information. The results cannot be easily interpreted with the exception of those concerning vitamin E, which seemed not related to lung cancer risk. In the ATBC-trial [10] and among smokers in a Finnish cohort study [6] no associations were observed between vitamin E and lung cancer too. In the latter study an inverse relation was reported among non-smokers, which we could not compare with our results as only 14 incident lung cancer cases occurred among non-smokers.

The difficulty in interpreting the results on vegetables, fruits, ß-carotene and vitamin C may in part be due to the large differences in intake across time, i.e. time trends in mean intakes and changes in ranking of subjects. No formal information on the validity of the dietary data is available, although the observed time trends in dietary intake are consistent with the expected lower intake with aging and the higher availability of fruits in the Netherlands [20]. The aggregate level of the questions on vegetable and fruit consumption probably has little effect on the quality of the data on total vegetable and fruit intake. In the Dutch food pattern many vegetables each contribute little to total vegetable intake, which may vary greatly across persons. Consequently, the data on ß-carotene, for which specific vegetables are the main source, might contain considerable measurement error. Assessed ß-carotene intake is therefore probably a better indicator of total vegetable intake ($r = 0.9$) than of true ß-carotene intake. We expect that the data on citrus fruits and other fruits represent the consumption of oranges and apples reasonably well, as these are by far the most commonly eaten fruits in the Netherlands. For this reason estimated vitamin C intake will probably contain less measurement error than ß-carotene intake.

There are three more points relevant to the interpretation of our results. First, subjects with different levels of vegetable, fruit, ß-carotene, and vitamin C intake differed in packyears of cigarettes smoked. Similar observations were reported for other populations including a Dutch one [21]. Residual confounding due to smoking, may therefore explain part of the associations with lung cancer. However, the results remained essentially the same when adjustment for smoking was done with five dummy variables instead of one continuous variable. Further, selection bias may have influenced our results. For the present analysis we excluded subjects for which no

dietary information was available in 1965 or 1970 as well as all cancer cases before 1970. Excluded subjects were older, had smoked more, and consumed less vitamin E than selected subjects ($p < 0.05$) and included 39 lung cancer cases. Selective loss of high-risk or more susceptible individuals may have occurred which would reduce the chance of finding inverse relationships. Such a selection bias would also explain the difference between the present results on fruits and vitamin C intake in 1960 and those of a previous analyses in which inverse relations were observed with 25-year lung cancer mortality [22]. Thirdly, we had planned to look at intake in single years and their average in relation to lung cancer risk. No a priori hypothesis was postulated concerning stable intake. It was not until the repeated dietary data were studied that the use of average intake as the only approach to combine the repeated measurements seemed inappropriate. This implies that our results on stable intake need to be replicated before firm conclusions can be drawn.

In the present study, intake of ß-carotene nor vegetables in any of the single years was associated with lung cancer risk. Vitamin C and fruit intake in 1970 were inversely related to lung cancer and results for vitamin C intake in 1960 pointed in the same direction. However, fruit intake in 1960 and 1965, and vitamin C intake in 1965 were not related to lung cancer risk. The presence of undiagnosed lung cancer which may have influenced intake of fruits and vitamin C in 1970 seems not the reason for the observed two-fold increased risks, as the results did not change substantially if cases diagnosed in the first five years of follow-up were excluded. The possibility that only recent intake is important is also not consistent with the above observation nor with the finding that the relative risks did not increase linearly with time.

If random measurement error was the reason why true inverse relationships were obscured in the analyses using a single intake measurement, then the relative risks using average intake (over 1960, 1965, and 1970) are expected to be higher [23]. This is apparently not the case, as the relative risks for low average intakes of vegetables, fruits and antioxidant (pro)vitamins were close to the means of the relative risks for low intakes in the separate years. The relative risk of a low stable intake of vitamin C was slightly higher than that of a low average intake. In contrast to this, we observed more than two-fold elevated risks on lung cancer for men with low stable intakes of ß-carotene, vegetables, and fruits. When the cut-off levels were changed from the 33th percentiles to the medians, the relative risks for low stable intakes of vegetables and fruits became less strong, and that of ß-carotene disappeared (data not shown). This

indicates that only subjects with a truly low intake may exhibit an increased lung cancer risk.

The differences in relative risks for low stable and low average intakes of ß-carotene, vegetables, and fruits may have several reasons. Firstly, it is possible that intakes of ß-carotene, vegetables, and fruits actually need to be sufficiently high over a long period of time in order to protect against lung cancer. In the second place, differences in average intakes between the subgroups with low and those with high stable intakes were about 50 percent larger than between the groups with low and high average intakes. Given a true underlying dose-response relationship, larger contrasts in exposure would result in higher relative risks, easier reaching statistical significance. Thirdly, the subjects with stable intakes might be selections of those subjects with less measurement error. If this were the case, relative risks for low stable intakes would have been less attenuated than relative risks for low average intakes. All three possibilities mentioned above support the hypothesis that low intakes of vegetables, fruits, and ß-carotene are associated with a higher risk of lung cancer.

Since no prospective studies are published in which repeated intake data of vegetables, fruits, or antioxidant (pro)vitamins are related to lung cancer risk, we can only compare our results with studies in which a single dietary assessment was used. One out of five cohort studies on total vegetable intake and lung cancer observed a statistically significant inverse association among women [24]. Of the four other studies, the results tended in the direction of an inverse association in a Norwegian study [25], among women in a retirement community in California [26], and among Finnish non-smokers [6], but not among the men of the Californian retirement community [26], Finnish smokers [6] and in the US Lutheran Brotherhood cohort [27]. For fruit intake, three out of eight cohort studies, including the previous analysis of this dataset, showed an inverse relationship [22,28] although one among non-smokers only [6]. Of five other studies on the effect of fruits, three tended in that direction [24,27,29], one did so for women only [26], and one did for squamous and small-cell carcinomas only [25].

Vitamin C intake was significantly inversely related to lung cancer mortality in a previous analysis of this dataset [22] and to lung cancer incidence among non-smokers, but not smokers, in a Finnish cohort study [6]. Of the four other cohort studies, the results of two tended in the direction of an inverse association [25,27], one did so for women, but not men [26], and one did when vitamin C from supplements

was not considered [24]. Similarly, intake of ß-carotene was inversely related to lung cancer in the first cohort study on this topic [30] and suggested an inverse association in three later ones [22,24,27], among women, but not men, in a California retirement community [26] and among non-smokers only [31]. Beta-carotene supplementation for 6-8 years among heavy smoking men in the ATBC-trial suggested adverse effects on lung cancer risk [10].

In conclusion, like other investigations the present study does not support the hypothesis that intake of vitamin E protects against lung cancer. For vitamin C intake our results, as well as those of most other cohort studies, point into the direction of an inverse association, although not entirely consistently. Our findings that low stable intakes of vegetables, fruits, and ß-carotene are associated with increased lung cancer risks while low average intakes are not, need to be replicated before conclusions can be drawn. The use of repeated intake measurements to select subgroups with stable, highly contrasting intakes may be a promising approach for studying diet-cancer relationships.

Acknowledgements

We would like to thank the fieldwork team in Zutphen, especially Dr. E.B. Bosschieter and A.M. Thomassen-Gijsbers, and Dr. M. Drijver, Dr. S.O. Keli, and I. Miedema for coding the morbidity data.

References

1. Visser O, Coebergh JWW, Schouten LJ (eds). Incidence of cancer in the Netherlands 1991. Utrecht: Netherlands Cancer Registry, 1994
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;66:1192-1308
3. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer. I. Epidemiology. Cancer Causes Control 1991;2:325-57
4. Jansen MCJF, Van 't Veer P, Kok FJ. Fruits and vegetables in chronic disease prevention. Rationale for fruits and vegetables-campaign. Report Department of Human Epidemiology and Health Science. Wageningen: Agricultural University, 1995
5. Peto R, Doll R, Buckley JD, et al. Can dietary beta-carotene materially reduce human cancer rates? Nature 1981;290:201-8
6. Knekt P, Järvinen R, Seppänen R, et al. Dietary antioxidants and the risk of lung cancer. Am J Epidemiol 1991;134:471-9
7. Comstock GW, Bush TL, Helzlsouer K. Serum retinol, beta-carotene, vitamin E, and selenium as related to subsequent cancer of specific sites. Am J Epidemiol 1992;135:115-21

8. Stähelin HB, Gey KF, Eichholzer M, et al. Plasma antioxidant vitamins and subsequent cancer mortality in the 12-year follow-up of the Prospective Basel Study. Am J Epidemiol 1991;133:766-75
9. Blot WJ, Li J-Y, Taylor PR, Li B. Lung cancer and vitamin supplementation (letter). N Engl J Med 1994;331:614
10. The Alpha-tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N Engl J Med 1994;330:1029-35
11. Hennekens CH, Buring JE, Peto R. Antioxidant vitamins - benefits not yet proved. N Engl J Med 1994;330:1080-1
12. Pryor WA. Lung cancer and vitamin supplementation (letter). N Engl J Med 1994;331:612
13. Hebert JR, Miller DR. Methodologic considerations for investigating the diet-cancer link. Am J Clin Nutr 1988;47:1068-77
14. Kromhout D, de Lezenne Coulander C, Obermann-de Boer GL, et al. Changes in food and nutrient intake in middle-aged men from 1960 to 1985 (the Zutphen Study). Am J Clin Nutr 1990;51:123-9
15. Feskens EJM, Weijenberg MP, Kromhout D. A longitudinal study on diet, risk factors and cardiovascular disease in an aging cohort: the Zutphen Study. Neth J Cardiol 1993;4:200-4
16. Hautvast JGAJ. Committee on uniform coding of dietary questionnaires: development of a system for computerized processing of dietary questionnaires (in Dutch). Voeding 1975;36:356-60
17. Commissie UCV. UCV table: extended food table 1985 (in Dutch). The Hague: Voorlichtingsbureau voor de Voeding, 1985
18. Vollebregt YCJ, Feskens EJM. Composing food tables with contents of retinol, ß-carotene, vitamin E and pectin for a.o. the Zutphen Study (in Dutch). Report no 441111002. Bilthoven: RIVM, 1993
19. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: John Wiley & Sons, 1980
20. FAO. Food balance sheets. 1984-1986 averages. Rome: FAO, 1991
21. Zondervan KT, Ocké MC, Smit HA, et al. Do dietary and supplementary intakes of antioxidants differ with smoking status? Int J Epidemiol (in press)
22. Kromhout D. Essential micronutrients in relation to carcinogenesis. Am J Clin Nutr 1987;45:1361-7
23. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press, 1992
24. Steinmetz KA, Potter JD, Folsom AR. Vegetables, fruit, and lung cancer in the Iowa Women's Health Study. Cancer Res 1993;53:536-43
25. Kvåle G, Bjelke E, Gart JJ. Dietary habits and lung cancer risk. Int J Cancer 1983;31:397-405
26. Shibata A, Paganini-Hill A, Ross RK, et al. Intake of vegetables, fruits, beta-carotene, vitamin C and vitamin supplements and cancer incidence among the elderly: a prospective study. Br J Cancer 1992;66:673-9
27. Chow W-H, Schuman LM, McLaughlin JK, et al. A cohort study of tobacco use, diet, occupation, and lung cancer mortality. Cancer Causes Control 1992;3:247-54
28. Fraser GE, Beeson WL, Phillips RL. Diet and lung cancer in California seventh-day adventists. Am J Epidemiol 1991;133:683-93

29. Long-de W, Hammond EC. Lung cancer, fruit, green salad and vitamin pills. Chin Med J 1985;98:206-10
30. Shekelle RB, Liu S, Raynor WJ jr, et al. Dietary vitamin A and risk of cancer in the Western Electric Study. Lancet 1981;ii:1185-90
31. Knekt P. Vitamin E and smoking and the risk of lung cancer. Ann NY Acad Sci 1993;280-7.

Chapter 4

Stability of blood (pro)vitamins during four years of storage at -20 °C. Consequences for epidemiologic research

Marga Ocké, Jaap Schrijver, Gatske Obermann-de Boer,
Bennie Bloemberg, Guido Haenen, Daan Kromhout

J Clin Epidemiol 1995; 48: 1077-85

Abstract

We studied the effects of frozen storage on (pro)vitamin concentrations in EDTA-plasma and whole blood. Aliquots from 55 samples were analyzed before storage and after 3, 6, 12, 24, 36 and 48 months at -20 °C. Dramatic decreases occurred for EDTA-plasma concentrations of vitamin E between 6 and 12 months, vitamin A, total carotenoids and ß-carotene after 1 year, and whole blood niacin. A smaller decrease was observed for folic acid at 1 year of storage, but the level remained constant thereafter. The vitamins D, B_6, B_{12} (EDTA-plasma), B_1 and B_2 (whole blood) showed no decline during 4 years of storage. With the exception of folic acid, the observed decreases varied considerably among subjects. Therefore using EDTA-plasma stored longer than 1 year at -20 °C will result in highly attenuated odds ratios when assessing the relationship between vitamin A, carotenoids, or vitamin E with a given disease. Attenuation will also occur when using niacin concentrations in whole blood stored for 4 years at -20 °C.

Introduction

Associations between concentrations of (pro)vitamins in serum or plasma and the risk of diseases, such as cancer and cardiovascular diseases are commonly studied [e.g. 1-6]. The study design especially suited for this purpose is the nested case-control study, since this type of study provides methodologically stronger evidence than an ordinary case-control study, and is more cost efficient than a full cohort analysis. However, in nested case-control studies a long interval of storage time between sample collection and laboratory analyses is inevitable. It is therefore important to know the effects of storage on concentrations of (pro)vitamins in serum, plasma or whole blood and its consequences for epidemiological research. Both effects on mean vitamin levels as well as effects on the correct ranking of individuals are important in this respect.

In a recent review on the stability of retinol, ß-carotene, and α-tocopherol, it was concluded that there were no studies available that met the criteria needed to ascertain with some degree of certainty the effects of frozen storage on the (pro)vitamin concentrations in serum or plasma. Those criteria were to assay fresh aliquots and aliquots stored for various periods of time, of the same plasma or serum sample, and to have enough of these comparisons to yield statistical precision [7]. Data for other vitamins is even more scarce. We therefore conducted a stability study on vitamin A, ß-carotene, total carotenoids, vitamin B_6, vitamin B_{12}, folic acid, vitamin D, and vitamin E in plasma with ethylenediaminetetraacetate (EDTA) as an anticoagulant, and vitamins B_1 and B_2 and niacin in whole blood stored at -20 °C. We describe here our observations on the stability of these (pro)vitamins when stored for up to 4 years and the consequences of the observed effects for epidemiological research.

Materials and methods

Blood and plasma samples

In 1988, fresh blood was obtained from 55 blood donors (male and female whites in the age range 20-55 years) through the Red Cross Blood Bank Foundation in Utrecht, located in the center of the Netherlands. About 450-500 ml of whole blood was collected in a standard blood bag to which 10 ml of an aqueous solution containing 750 mg (1.86 mmol) K_2H_2EDTA was added as anticoagulant. The quality and safety of the collected blood as well as the apparent health of the donor were checked by the standard procedures of the blood bank. Blood bags of at least four people were

collected and processed each week, resulting in a total collection period of about 3 months for blood from 55 people.

All donor blood was processed the day after it was collected. During temporary storage and transport the temperature was kept at about +4 °C. At the TNO-Nutrition Institutes Laboratory (Zeist, The Netherlands), blood from each donor was carefully homogenized. About 100 aliquots of 1.5 ml of whole blood were collected in 3.6 ml cryotubes (Nunc, Roskilde, Denmark) and stored at -20 °C in the freezer of the TNO laboratory. The remaining whole blood from each donor (about 350 ml) was divided in 10 ml aliquots and centrifuged for 10 min at 1500 g at room temperature. Thereafter, plasma was separated and stored as 1.5 ml aliquots in 3.6 ml cryotubes at -20 °C. Whole blood and plasma samples were transported within 1 week on solid carbon dioxide (-76 °C) from the TNO laboratory to a large freezer located in another city for long-term carefully controlled storage. The temperature of this freezer averaged -23 °C, and did not exceed -20 °C during the storage period.

A few aliquots of whole blood and plasma from each donor were not frozen and were used for analysis of the (pro)vitamins of interest on the day the blood was processed, which means within 24-48 hr after collection. Analytical values obtained in this way served as fresh sample levels. After 3, 6, 12, and 24 months of storage, the analyses of all vitamins were repeated. For vitamin A, ß-carotene, total carotenoids and vitamin E the analyses were then canceled, since these vitamin levels declined considerably. For the other vitamins two more series of analyses were performed after 36 and 48 months of storage. As a result of the 3 months needed for collection of all 55 blood bags, and since laboratory analyses were performed simultaneously, the storage time at the time of analyses differed across the donors.

Analytical procedures

Aliquots of whole blood and plasma from each donor were collected from the bulk freezer, transported on solid carbon dioxide to the TNO laboratory and analyzed for the various compounds within a few days. Temporary storage at the laboratory was done at -20 °C. All manipulations with blood and plasma samples were carried out under subdued light conditions. On the day of analysis frozen samples were thawed in running tap water (about +15 °C) and extracted immediately thereafter.

Whole blood was used for high-performance liquid chromatographic (HPLC) analysis of vitamin B_1 as total thiamin [8], and vitamin B_2 as its cofactor flavin adenine

dinucleotide [9], and for microbiological determination of niacin (nicotinic acid plus nicotinamide) using *Lactobacillus plantarum* (ATCC 8014) as the test organism [10]. All other compounds were determined in EDTA-plasma. Vitamin A (all-*trans* retinol), total carotenoids, ß-carotene, and vitamin E (α-tocopherol) were quantified by HPLC with variable ultraviolet detection. Detection after separation on a Hypersil column was carried out using one detector set at 350 nm (retinol) and a second detector switching between 445 (ß-carotene) and 292 nm (α-tocopherol) [11]. Vitamin D was analyzed as 25-hydroxyvitamin D by a competitive protein binding assay after organic extraction and chromatographic sample clean-up to remove other vitamin D compounds [12]. Vitamin B_6 was determined as its cofactor pyridoxal 5'-phosphate using a radioenzymatic method with apoenzyme of tyrosine carboxylase [13]. Vitamin B_{12} ('true' cobalamin) and folic acid (5-methyl tetrahydrofolic acid) were quantified by competitive protein binding assay using the SimulTrac radioassay kit (Becton-Dickenson Immunodiagnostics, Orangeburg, NY, USA) [14].

All samples for all vitamins were analyzed in duplicate. Out of the 69 series of analyses, 47 had a within-duplo coefficient of variation below 5%, 18 between 5 and 10%, and four above 10%. Three of these four high coefficients of variation concern the analyses of total carotenoids, ß-carotene and vitamin E at 2 years of storage, with highly declined levels of vitamins, the other was for the analyses of ß-carotene at 3 months of storage.

In each series of analyses at least three aliquots in duplicate of quality control samples of whole blood or serum were incorporated. These control samples were stored in a controlled freezer at -20 °C. After a maximum of 6 months a new pool of quality control samples was used. The analytical results obtained for the control samples, randomly included in each series, were used to judge the quality of the analytical results of the unknown test samples. Results were discarded and analyses repeated within 1 week if the concentrations found for the quality control samples in a series differed significantly from the theoretical values at the level of $p < 0.025$.

Statistic evaluation

The average value of the duplicate measurements was used as the vitamin level for the statistical evaluation described below. Means and standard deviations were calculated for the vitamin levels in the fresh samples and in the samples stored at -20 °C for on average 3, 6, 12, 24, and (if available) 36 and 48 months. Two-tailed paired *t*-tests

were performed to test whether the concentrations after storage differed significantly from the fresh sample concentrations ($p < 0.025$).

Pearson product-moment correlation coefficients were calculated between the fresh sample vitamin levels and the levels after storage. In addition to this, partial correlation coefficients adjusted for exact storage time were calculated. Although the distributions of ß-carotene, vitamins B_6 and B_{12}, and folic acid were not normal, Spearman correlation coefficients did not materially differ from Pearson correlation coefficients. Storage effects on the classification of vitamin levels into tertiles were also evaluated. Tertiles instead of quartiles or quintiles were chosen because of the relatively small numbers.

Simulation study

A simulation was carried out to illustrate the consequences of using vitamin levels in stored samples instead of fresh samples for estimating odds ratios for a given disease. Four of the vitamins in our stability study which differed in stability were chosen as examples: total carotenoids, folic acid, vitamin B_1, and niacin. For the simulation, a case-control study with 270 cases and 270 controls was assumed. Of the cases 110 subjects were assumed to be in the first tertile of the fresh sample vitamin level, 90 in the second, and 70 in the third. For the controls it was assumed the other way around: 70, 90 and 110 subjects in the first, second and third tertile, respectively, of the fresh sample vitamin level. Thus, it was assumed that subjects in the third tertile of fresh vitamin level compared to those in the first tertile have an odds ratio with a Taylor-series 95% confidence interval (CI) [15] of 0.40 (CI 0.27-0.62), while subjects in the second tertile compared to those in the first tertile have an odds ratio of 0.64 (CI 0.42-0.97).

The odds ratios and their confidence intervals which would be observed when vitamin levels in samples stored at -20 °C would have been used instead of levels in fresh samples, are calculated by adjusting the numbers of cases and controls in the three tertiles of vitamin levels for the misclassification observed in our stability study. The adjusted numbers of cases (d_i) and controls (w_i) in the i-th tertile of vitamin level are:

$$d_i = 110 * p_{i1} + 90 * p_{i2} + 70 * p_{i3}$$
$$w_i = 70 * p_{i1} + 90 * p_{i2} + 110 * p_{i3}$$

where p_{ij} is taken from our stability study, and is the proportion of the fresh samples in tertile j, that is in tertile i after storage at -20 °C. The numbers in the above equations are the assumed true numbers of cases and controls for vitamin tertiles. The equations are derived from Willett [16] who illustrates attenuation of odds ratios with two levels of exposure.

Results

At the end of the 3-month collection period, blood was obtained from 32 males and 23 females, with an average age (± standard deviation) of 38.8 ± 8.2 and 29.7 ± 9.5 years, respectively. The mean values and standard deviations for the fat-soluble (pro)vitamins, and for the water-soluble vitamins across storage periods are presented in tables 1 and 2, respectively. Distributions of fresh sample values are similar to the reference ranges observed for Dutch healthy adults [17]. At 1 year of storage decreases of 20-30% were found for vitamin A, ß-carotene and total carotenoids, while 2 years at -20 °C resulted in a decline of two thirds or more. For vitamin E larger decreases were observed: at 1 year of storage only about half of the fresh sample level was recovered, while after 2 years the average level dropped to < 10% of the fresh sample mean level. The mean level of niacin was about 25% lower at 4 years of storage compared to the fresh samples. The average folic acid level at 1 year of storage was about 20% lower than the initial level, and remained about the same thereafter. As for the other vitamins significant differences from the fresh sample values occurred in both directions, but no decline of more than 20% was observed.

Pearson correlation coefficients (tables 1 and 2) were below 0.5 for vitamin E at 1 year or more of storage, for vitamin A, ß-carotene, and total carotenoids at 2 years of storage, and for niacin at 3 and 4 years of storage. Most of these correlation coefficients were higher if adjusted for exact storage time, but they remained at 0.5 or lower. For the other vitamins, including folic acid, the correlation coefficients remained high: the majority above 0.8. The percentages in the same and opposite tertiles in fresh samples compared to the samples after storage showed the same pattern as the correlation coefficients.

Table 1. *Fat-soluble (pro)vitamin concentrations in EDTA-plasma of 55 Dutch blood donors during frozen storage at -20 °C*

Vitamin	Storage time (months)	Mean	sd	Pearson r[a]	Pearson r[a] adjusted[b]	% in same tertile[a]	% in opposite tertile[a]
Vitamin A	0	1.85	0.40				
(µmol/l)	3	1.59*	0.38	0.93	0.94	76	0
	6	1.73*	0.41	0.92	0.92	63	0
	12	1.40*	0.42	0.73	0.78	59	2
	24	0.66*	0.53	0.04	0.16	30	24
Beta-carotene	0	0.30	0.17				
(µmol/l)	3	0.27*	0.17	0.95	0.96	85	0
	6	0.27*	0.17	0.92	0.95	65	0
	12	0.22*	0.14	0.91	0.93	76	0
	24	0.04*	0.05	0.24	0.34	39	28
Total	0	1.72	0.64				
carotenoids	3	1.72	0.70	0.93	0.95	78	0
(µmol/l)	6	1.64*	0.66	0.94	0.95	85	0
	12	1.22*	0.58	0.80	0.83	83	2
	24	0.47*	0.36	0.30	0.35	39	15
Vitamin D	0	73	23				
(nmol/l)	3	72	22	0.90	0.90	58	0
	6	56*	18	0.84	0.83	64	2
	12	69*	20	0.86	0.85	67	0
	24	81*	23	0.90	0.89	73	0
	36	73	21	0.82	0.81	69	4
	48	63*	20	0.83	0.81	69	4
Vitamin E	0	25.0	6.1				
(µmol/l)	3	24.0*	6.1	0.91	0.96	67	0
	6	23.3*	5.4	0.86	0.88	78	2
	12	13.0*	7.3	0.21	0.37	38	13
	24	1.8*	2.9	-0.15	-0.06	33	27

* $p < 0.025$ for paired t-test comparing sample values after storage with the fresh sample values
[a] compared to fresh sample levels; [b] partial correlation coefficients adjusted for exact storage time

Table 2. *Water-soluble vitamin concentrations in EDTA-plasma or whole blood of 55 Dutch blood donors during frozen storage at -20 °C*

Vitamin	Storage time (months)	Mean	sd	Pearson r[a]	Pearson r[a] adjusted[b]	% in same tertile[a]	% in opposite tertile[a]
Vitamin B_1	0	128	23				
(nmol/l	3	126	22	0.92	0.92	75	0
whole blood)	6	123*	21	0.91	0.92	75	0
	12	125*	22	0.91	0.92	73	0
	24	127	22	0.89	0.90	67	0
	36	132*	23	0.87	0.89	69	0
	48	127	23	0.85	0.88	64	4
Vitamin B_2	0	0.27	0.03				
(µmol/l	3	0.30*	0.04	0.83	0.83	71	4
whole blood)	6	0.29*	0.04	0.91	0.91	76	2
	12	0.28	0.04	0.73	0.72	73	4
	24	0.27*	0.04	0.88	0.87	69	4
	36	0.30*	0.04	0.85	0.84	64	2
	48	0.28	0.03	0.91	0.90	80	0
Vitamin B_6	0	41	17				
(nmol/l	3	45*	21	0.96	0.96	80	0
EDTA-plasma)	6	43*	18	0.95	0.96	82	0
	12	40	18	0.96	0.96	75	0
	24	43	21	0.94	0.94	82	0
	36	45*	22	0.92	0.93	78	0
	48	36*	20	0.94	0.95	78	0
Vitamin B_{12}	0	260	100				
(pmol/l	3	261	102	0.99	0.99	93	0
EDTA-plasma)	6	255	95	0.98	0.99	93	0
	12	264	113	0.99	0.99	93	0
	24	254*	89	0.99	0.99	93	0
	36	255	94	0.98	0.98	93	0
	48	240*	101	0.99	0.99	93	0
Folic acid	0	10.7	3.9				
(nmol/l	3	11.7*	4.6	0.97	0.96	85	0
EDTA-plasma)	6	9.9*	3.3	0.93	0.93	78	0
	12	8.7*	2.8	0.91	0.91	78	0
	24	8.8*	2.7	0.90	0.90	67	0
	36	9.0*	2.8	0.89	0.90	71	0
	48	8.3*	2.6	0.89	0.90	80	0
Niacin	0	46	5.4				
(µmol/l	3	45*	5.7	0.71	0.69	60	0
whole blood)	6	49*	5.9	0.69	0.73	60	5
	12	46	4.9	0.68	0.71	55	5
	24	47	6.5	0.56	0.55	64	5
	36	45	7.0	0.44	0.50	55	9
	48	36*	11.2	0.39	0.38	44	13

* $p < 0.025$ for paired t-test comparing sample values after storage with the fresh sample values
[a] compared to fresh sample levels; [b] partial correlation coefficients adjusted for exact storage time

The simulated effects on odds ratios of using total carotenoid, folic acid, vitamin B_1, and niacin levels in stored samples instead of fresh samples, are given in table 3. For total carotenoids, the 'true' odds ratios of 0.64 and 0.40 for the second and third tertile were observed as 1.10 and 0.83, and were no longer significant at the $p = 0.05$ level, when using vitamin levels in plasma stored for 2 years. Shorter storage did not have such large effects. The observed odds ratios for folic acid hardly differed from the assumed 'true' odds ratios. For vitamin B_1 the observed odds ratios increased towards 1 with increasing storage time, but when whole blood stored for 4 years was used the observed odds ratio for the third tertile was still 0.56. The observed odds ratios for niacin also inclined towards 1 with longer storage. When using blood stored for 4 years, the odds ratios were 0.95 and 0.77 for the second and third tertile, respectively.

Discussion

The main results of the present study showed that storage of EDTA-plasma at -20 °C caused a dramatic decline in vitamin E between 6 and 12 months, and in vitamin A, total carotenoids and ß-carotene after about 1 year, while no large decline occurred for vitamin D, vitamin B_6, and vitamin B_{12} for up to 4 years of storage. Folic acid levels were about 20% lower at 1 year of storage and remained about the same thereafter. In whole blood with EDTA as anticoagulant, vitamins B_1 and B_2 were stable up to 4 years, but niacin showed a significant decline during the storage period. With the exception of folic acid, the observed decreases differed largely across subjects, since correlation coefficients between stored and fresh vitamin levels decreased simultaneously with decreases in mean vitamin levels.

Laboratory analyses

The imprecision of the analytical method may complicate the interpretation of the results obtained. For some of the vitamins the mean value showed an increase, a systematic error is likely. Folic acid is an example in which the changes in the mean level could either be caused by real decline or by between-run imprecision of the analytical method. Due to the use of quality control samples, large systematic errors are excluded.

Table 3. *Simulated effects of using samples stored at -20 °C on observed odds ratios (95% confidence intervals)*

Vitamin	Storage time (months)	Tertile		
		1	2	3
Total carotenoids in EDTA-plasma	0[a]	1.00	0.64 (0.42-0.97)	0.40 (0.27-0.62)
	3	1.00	0.74 (0.49-1.12)	0.47 (0.31-0.72)
	6	1.00	0.67 (0.44-1.02)	0.45 (0.29-0.68)
	12	1.00	0.74 (0.49-1.12)	0.47 (0.31-0.72)
	24	1.00	1.10 (0.73-1.65)	0.83 (0.55-1.27)
Folic acid in EDTA-plasma	0[a]	1.00	0.64 (0.42-0.97)	0.40 (0.27-0.62)
	3	1.00	0.62 (0.41-0.94)	0.45 (0.29-0.68)
	6	1.00	0.64 (0.42-0.97)	0.47 (0.31-0.72)
	12	1.00	0.64 (0.42-0.97)	0.47 (0.31-0.72)
	24	1.00	0.69 (0.45-1.04)	0.51 (0.33-0.77)
	36	1.00	0.65 (0.43-0.99)	0.49 (0.32-0.75)
	48	1.00	0.64 (0.42-0.96)	0.46 (0.30-0.71)
Vitamin B_1 in whole blood	0[a]	1.00	0.64 (0.42-0.97)	0.40 (0.27-0.62)
	3	1.00	0.77 (0.51-1.17)	0.48 (0.32-0.73)
	6	1.00	0.72 (0.47-1.09)	0.48 (0.32-0.73)
	12	1.00	0.78 (0.51-1.18)	0.49 (0.32-0.75)
	24	1.00	0.79 (0.52-1.20)	0.51 (0.33-0.77)
	36	1.00	0.82 (0.54-1.24)	0.50 (0.33-0.76)
	48	1.00	0.68 (0.45-1.03)	0.56 (0.37-0.86)
Niacin in whole blood	0[a]	1.00	0.64 (0.42-0.97)	0.40 (0.27-0.62)
	3	1.00	0.71 (0.47-1.06)	0.53 (0.35-0.82)
	6	1.00	0.81 (0.54-1.22)	0.59 (0.38-0.91)
	12	1.00	0.83 (0.55-1.27)	0.62 (0.41-0.93)
	24	1.00	0.69 (0.46-1.05)	0.58 (0.38-0.88)
	36	1.00	0.80 (0.52-1.20)	0.66 (0.43-1.01)
	48	1.00	0.95 (0.63-1.44)	0.77 (0.51-1.16)

[a] assumed true odds ratios and confidence intervals

Random measurement error in the vitamin analyses is not likely to be a large problem either, as most of the within-run precision coefficients of variation are below 5%. This is also indicated by the high correlations between the fresh sample vitamin

levels and those after 3 months of storage for all vitamins, except for niacin which has a correlation coefficient of 0.71. The correlation coefficients after longer storage for niacin show therefore a combined effect of random measurement error in the vitamin analyses and between-subject differences in instability.

The variation in storage period at the time of analyses resulted in some random error for the unstable analytes, as was shown in the results section. This did not however lead to other conclusions.

Comparison with other stability studies

In this study, the effects of storage time, up to 4 years at -20 °C, are reported for 11 vitamins and provitamins based on blood from 55 people. To our knowledge the present study is more comprehensive than other stability studies reported in the literature so far. Unfortunately, as a result of logistic problems we could not obtain blood samples for long-term storage and analyses of vitamin C. Our results do not include the effects of repeated freeze-thaw cycles, which may also effect stability [18].

In contrast to our results, other stability studies on vitamin A have found no large decline in mean values in serum or plasma stored at -20 °C [7]. One study reported stability in serum for as long as 8 years [19]. The between-subject differences in decline in our study are also larger than those found in other studies [3,19]. One possible explanation for the relatively short stability we observed, is the addition of EDTA as anticoagulant which was not used in other studies. Two studies on the effects of different anticoagulants showed small [20] and large [21] decreases in retinol concentrations for EDTA, possibly due to a degradation product which was formed [21]. Plasma anticoagulated with heparin demonstrated retinol values equivalent to serum, while potassium oxalate and sodium citrate caused 20% lower values [20].

Other stability studies on carotenoids, although often with small or unstated numbers, suggest that important losses of carotenoids occur between 5 and 15 months of storage at -20 °C [7]. Our results fit well with this picture. Three studies on the effects of storage at -70 °C did not observe a decline in mean carotenoid levels [22-24]. The longest storage period studied was 5 years [24].

Vitamin E in serum or plasma was reported to be stable at -20 °C for up to 15/16 months by two studies [23,25] while another study reported serious degradation in serum stored for 7-13 months [18]. Our results are in line with this last study. A comparison of the mean α-tocopherol concentrations in serum or plasma from control

subjects in case-control studies led to the conclusion that some losses can be expected at temperatures above -40 °C [7]. A reliability coefficient of 0.65 was observed between serum stored for 4 years at -20 °C and fresh sample levels [26], which is high compared to our results.

Information on the stability of vitamin D and the water-soluble vitamins is scarce. One study reported that vitamin D in plasma seems stable for up to 22.5 months, but it was not specified at which temperature the samples were frozen [27]. Samples stored for 11 months at -18 °C showed 10% lower vitamin D levels than samples stored for 1 month [28]. Our results also showed differences in mean vitamin D levels of about 10% at some storage times but this seems to be due to systematic differences in the laboratory measurements, since an increase of about 10% also occurred.

For plasma pyridoxal 5'-phosphate, the vitamin B_6 cofactor, similar levels in samples stored for 1-2 years at -30 °C and in fresh samples were observed, while the correlation coefficient was 0.95 [29]. This is in accordance with our results at -20 °C. Another study reported a decline of about 2.2% per year during a 700 day storage at -20 °C [30].

For vitamin B_1, vitamin B_2, niacin, folic acid, and vitamin B_{12} no stability data could be found in the literature, although it is stated that vitamin B_1 [8] and vitamin B_2 [9] in whole blood are stable for several months at -20 °C, niacin for much longer [10], while serum or plasma vitamin B_{12} can be stored for at least 1 year at -20 °C without a serious decline [14]. Our study reports stability for up to 4 years for these vitamins which is as long or longer than stated.

Consequences for epidemiological studies

Due to the already known instability of carotenoids and vitamin E at -20 °C [7], storage at very low temperatures like -70, -80 or even -196 °C is recommended and nowadays practiced [31,32]. In prospective studies with many thousands of subjects and long follow-up times the costs of storage at very low temperatures are enormous. These costs could be reduced substantially when serum, plasma or whole blood for the analyses of those vitamins that are more stable could be stored at higher temperatures. Vitamins B_1 and B_2 in whole blood, and folic acid, vitamins B_6, B_{12} and D in EDTA-plasma are possible candidates, but data for longer storage are needed to make firm conclusions.

To discuss the consequences of the instability of (pro)vitamins during freezer storage on epidemiological measures of effect, systematic decline as well as the decline that differs across subjects should be considered. We have shown that decreases in mean levels and in correlations with fresh sample levels occur simultaneously for vitamin A, total carotenoids, ß-carotene, and vitamin E in EDTA-plasma, and niacin in whole blood with EDTA as anticoagulant. Why the rate of decline varied so much among subjects is not known. It may be a function of other compounds in the plasma or whole blood, eg lipid peroxides or antioxidants. As long as this is not known the between-subject differences in decline can be treated as random effects. For folic acid the decrease was mainly systematic, since the correlation coefficients with the fresh sample values remained high.

The systematic decline that equally occurs in cases and controls does not affect measures of effect [16], as is illustrated in our simulation with the folic acid stability data. However, no conclusions with respect to absolute vitamin levels can be made. Furthermore, systematic decline can lead to serious differential systematic bias and therefore incorrect associations when differences in handling specimens between cases and controls occur. This was illustrated by Wald *et al.* [33]. They concluded that their results, i.e. that women who subsequently developed breast cancer had significantly lower vitamin E levels than matched controls [34], may have been artefactual due to systematic differences in decline in vitamin E levels between cases and controls upon handling and storage. Matching cases and controls on the date of blood taken, carefully handling their samples equally, and analyzing their samples in the same run could have prevented differential systematic decline.

Random between-subject differences in decline tend to attenuate epidemiological measures of effect towards the no association value [16], as is also shown in our simulation with the total carotenoid and niacin stability data. This simulation merely had the object to illustrate the consequences of (pro)vitamin instability during storage on odds ratios calculated from vitamin levels in stored samples. The number of samples in our stability study however is too small to make precise estimates of these effects. It is recommended that a stability study is incorporated into the study design of each nested case-control study to be able to correct the epidemiological measures of effect for attenuation.

Conclusion

In conclusion, this study demonstrates that EDTA-plasma stored for longer than 1 year at -20 °C should not be used to assess the relationship between vitamin A, carotenoids or vitamin E and a given disease or any other variable of interest, since measures of effect will be considerably attenuated due to between-subject differences in decline in these vitamin levels. The same can be concluded for niacin in whole blood with EDTA as anticoagulant, stored for 4 years at -20 °C. For vitamin A, serum or heparinized plasma seems more suitable than EDTA-plasma. For the other unstable (pro)vitamins storage at lower temperatures is advisable. In the meantime more research is needed to show under which temperature and what conditions acceptable stability occurs. The vitamins B_6 and B_{12}, folic acid, and vitamin D in EDTA-plasma, and vitamins B_1 and B_2 in whole blood stored for 4 years at -20 °C, can be used to assess odds ratios for a given disease. Although for folic acid no conclusions can be drawn about the absolute vitamin levels. More research is needed to evaluate effects of longer storage at -20 °C.

Acknowledgements

Support for this study was received from the Ministry of Welfare, Public Health and Culture, The Netherlands. The authors thank the following persons for their assistance: Laboratory measures were made by Jan van Schoonhoven and Frank van Schaik, and Dr. Bas Bueno de Mesquita gave editorial suggestions. The Red Cross Blood Bank Foundation in Utrecht is gratefully acknowledged for supplying the blood.

References

1. Connett JE, Kuller LH, Kjelsberg MO, Polk BF, Collins G, Rider A, Hulley SB. Relationship between carotenoids and cancer. The multiple risk factor intervention trial (MRFIT) study. Cancer 1989; 64: 126-134
2. Kok FJ, Bruijn AM de, Vermeeren R, Hofman A, Laar A van, Bruin M de, Hermus RJJ, Valkenburg HA. Serum selenium, vitamin antioxidants, and cardiovascular mortality: a 9-year follow-up study in the Netherlands. Am J Clin Nutr 1987; 45: 462-468
3. Knekt P, Aromaa A, Maatela J, Aaran R-K, Nikkari T, Hakama M, Hakulinen T, Peto R, Teppo L. Serum vitamin A and subsequent risk of cancer: cancer incidence follow-up of the Finnish mobile clinic health examination survey. Am J Epidemiol 1990; 132: 857-870
4. Riemersma RA, Wood DA, Macintyre CCA, Elton RA, Gey KF, Oliver MF. Risk of angina pectoris and plasma concentrations of vitamins A, C, and E and carotene. Lancet 1991; 337: 1-5
5. Wald NJ, Thompson SG, Densem JW, Boreham J, Bailey A. Serum vitamin E and subsequent risk of cancer. Br J Cancer 1987; 56: 69-72

6. Willett WC, Polk BF, Underwood BA, Stampfer MJ, Pressel S, Rosner B, Taylor JO, Schneider K, Hames CG. Relation of serum vitamins A and E and carotenoids to the risk of cancer. N Engl J Med 1984; 310: 430-434
7. Comstock GW, Alberg AJ, Helzlsouer KJ. Reported effects of long-term freezer storage on concentrations of retinol, ß-carotene, and α-tocopherol in serum or plasma summarized. Clin Chem 1993; 39: 1075-1078
8. Schrijver J, Speek AJ, Berg H van den. Total thiamin in whole blood by HPLC. In: Fidanza F (Ed). Nutritional status assessment. A manual for population studies. London: Chapman & Hall; 1991: 235-241
9. Schrijver J, Speek AJ, Berg H van den. Flavin adenine dinucleotide (FAD) in whole blood by HPLC. In: Fidanza F (Ed). Nutritional status assessment. A manual for population studies. London: Chapman & Hall; 1991: 251-256
10. Schrijver J, Breederode N van, Berg H van den. Niacin in whole blood by microbiological assay. In: Fidanza F (Ed). Nutritional status assessment. A manual for population studies. London: Chapman & Hall; 1991: 258-263
11. Vliet T van, Schaik F van, Schoonhoven J van, Schrijver J. Determination of several retinoids, carotenoids and E vitamers by high performance liquid chromatography. Application to plasma and tissues of rats fed a diet rich in either ß-carotene or canthaxanthin. J Chromatogr 1991; 553: 179-186
12. Berg H van den, Schrijver J, Boshuis PG. Vitamin D (25-OHD) in serum by competitive protein-binding assay. In: Fidanza F (Ed). Nutritional status assessment. A manual for population studies. London: Chapman & Hall; 1991: 203-209
13. Chabner B, Livingston D. A simple enzymic assay for pyridoxal phosphate. Analyt Biochem 1970; 34: 413-423
14. Berg H van den, Schrijver J. Vitamin B_{12} (total cobalamins) in serum by competitive protein-binding assay. In: Fidanza F (Ed). Nutritional status assessment. A manual for population studies. London: Chapman & Hall; 1991: 290-296
15. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research. Principles and quantitative methods. London: Wadsworth; 1982: 299
16. Willett W. Nutritional epidemiology. New York: Oxford University Press; 1990: 272-291
17. Schrijver J, Veelen BWC van, Schreurs WHP. Biochemical evaluation of the vitamin and iron status of an apparently healthy Dutch free-living elderly population. Comparison with younger adults. Int J Vit Nutr Res 1985; 55: 337-349
18. Gunter EW, Driskell WJ, Yeager PR. Stability of vitamin E in long-term stored serum. Clin Chim Acta 1988; 175: 329-336
19. Driskell WJ, Lackey AD, Hewett JS, Bashor MM. Stability of vitamin A in frozen sera. Clin Chem 1985; 31: 871-872
20. Nierenberg DW. Determination of serum and plasma concentrations of retinol using high-performance liquid chromatography. J Chromatogr 1984; 311: 239-248
21. McClean SW, Ruddel ME, Gross EG, DeGiovanna JJ, Peck GL. Liquid-chromatographic assay for retinol (vitamin A) and retinol analogs in therapeutic trials. Clin Chem 1982; 28: 693-696
22. Mathews-Roth MM, Stampfer MJ. Some factors affecting determination of carotenoids in serum. Clin Chem 1984; 30: 459-461
23. Craft NE, Brown ED, Smith JC Jr. Effects of storage and handling conditions on concentrations of individual carotenoids, retinol, and tocopherol in plasma. Clin Chem 1988; 34: 44-48

24. Edmonds BK, Nierenberg DW. Serum concentrations of retinol, d-α-tocopherol and ß-carotene: effects of storage at -70 °C for five years. J Chromatogr 1993; 614: 169-174
25. Driskell WJ, Neese JW, Bryant CC, Bashor MM. Measurement of vitamin A and vitamin E in human serum by high-performance liquid chromatography. J Chromatogr 1982; 231: 439-444
26. Knekt P, Aromaa A, Maatela J, Aaran R-K, Nikkari T, Hakama M, Hakulinen T, Peto R, Saxén E, Teppo L. Serum vitamin E and risk of cancer among Finnish men during a 10-year follow-up. Am J Epidemiol 1988; 127: 28-41
27. Stamp TCB, Round JM. Seasonal changes in human plasma levels of 25-hydroxyvitamin D. Nature 1974; 247: 563-565
28. Norris RLG, Thomas MJ, Craswell PW. Assessment of a two-step high-performance liquid chromatographic assay using dual-wavelength ultraviolet monitoring for 25-hydroxyergocalciferol and 25-hydroxycholecalciferol in human serum or plasma. J Chromatogr 1986; 381: 53-61
29. Borschel MW, Kirksey A, Hamaker BR. A micromethod for determination of plasma pyridoxal phosphate and its use in assessment of storage stability of the vitamer. J Pediatr Gastroenterol Nutr 1987; 6: 409-413
30. Howard MP, Andon MA, Reynolds RD. Long-term stability of pyridoxal phosphate in frozen human plasma (abstract). Fed Proc 1984; 43: 486
31. Hunter D. Biochemical indicators of dietary intake. In: Willett W. Nutritional epidemiology. New York: Oxford University Press; 1990: 143-216
32. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol 1992; 3: 783-791
33. Wald NJ, Nicolaides-Bouman A, Hudson GA. Plasma retinol, beta-carotene and vitamin E levels in relation to the future risk of breast cancer (letter). Br J Cancer 1988; 57: 235
34. Wald NJ, Boreham J, Hayward JL, Bulbrook RD. Plasma retinol, ß-carotene and vitamin E levels in relation to the future risk of breast cancer. Br J Cancer 1984; 49: 321-324

Chapter 5

The Dutch EPIC food frequency questionnaire I Description of the questionnaire, and relative validity and reproducibility for food groups

Marga Ocké, Bas Bueno de Mesquita, Ester Goddijn, Annemarie Jansen, Margreet Pols, Wija van Staveren, Daan Kromhout

Int J Epidemiol, in press

Abstract

A self-administered food frequency questionnaire was developed for the Dutch cohort of the EPIC study. Habitual consumption of 178 food items can be calculated from the questionnaire data. Reproducibility and relative validity for food group intake were investigated in a population of 121 Dutch men and women. The questionnaire was administered three times at six-month intervals in order to determine the reproducibility. To assess the relative validity 12 monthly 24-h recalls served as reference method. Spearman correlation coefficients between estimates of food group intake assessed by repeated questionnaires ranged from 0.45-0.92. For men, Spearman correlation coefficients between estimates of food group intake based on the questionnaire and those based on 24-h recalls ranged from 0.21 for cooked vegetables to 0.78 for sugar & sweet products, with a median of 0.61. For women the median was 0.53, with a minimum of 0.31 for vegetables and a maximum of 0.87 for alcoholic beverages. The photographs in the questionnaire for the estimation of portion sizes contributed little to the relative validity of the ranking of subjects. However, on the group level most median food group estimates based on photographic portion sizes were closer to the median intakes as assessed by 24-h recalls than those based on standard portion sizes. In conclusion, the questionnaire seems adequate for ranking Dutch EPIC subjects according to intake of most food groups, although the relative validity for some food groups, such as vegetables and fish, remains of concern.

Introduction

For large-scale epidemiological studies on chronic diseases, food frequency questionnaires are often the method of choice to obtain dietary exposure data. The two main reasons for this choice are the aim to measure habitual long-term dietary intake and the fact that the method is relatively inexpensive since highly trained interviewers are not required [1]. Some of the recent food frequency questionnaires include questions on habitual portion size, but it has been questioned whether this improves the validity of the method [2,3].

The relative validity of food group intake estimated by semi-quantitative food frequency questionnaires is reported less often than that of nutrient intake. Knowledge about this aspect is however important since it indicates more directly those questions or items in the questionnaire that should be considered for improvement, and since many epidemiological studies report relative risks for different levels of food group intake rather than nutrient intake.

Within the context of the European Prospective Investigation into Cancer and Nutrition (EPIC) [4], we developed a semi-quantitative food frequency questionnaire to be used for the Dutch cohorts. The aim was to assess the intake of energy, macronutrients, dietary fibre, retinol, vitamins C and E, ß-carotene, and food groups considered to be important in cancer aetiology. Since analyses will be performed both within and between EPIC cohorts, knowledge about random and systematic measurement errors in the dietary intake estimates for the study populations is crucial [5].

During the pilot phase of the EPIC study in 1991/1992 we tested the newly developed food frequency questionnaire among 121 men and women. In the present paper the questionnaire and its development are described, and results for the food groups are reported. A companion paper in this issue covers energy and nutrients [6]. For the food groups 6 and 12-month reproducibility and the relative validity compared to 12 repeated 24-h dietary recalls are presented. Furthermore the question whether portion size information as provided by the food photographs included in the questionnaire improves the quality of the dietary data is evaluated.

Material and methods

The food frequency questionnaire

The self-administered food frequency questionnaire contains questions on the average consumption frequency during the past year for 79 main food items. Subjects can indicate their answers in times per day, per week, per month or per year, or as never. For several food items additional questions are asked about the consumption frequency for different subitems, preparation methods, or additions. For these questions four multiple choice categories are given, i.e. always/mostly, often, sometimes, and seldom/never. The definitions of these terms are given in the instructions to the questionnaire. This approach was chosen to avoid overestimation and internal inconsistency, which commonly occur with food frequency questionnaires [7]. In total, the habitual consumption of 178 food items can be calculated from the information thus obtained.

The food items in the questionnaire were selected by a data-based approach [8], using the Dutch National Food Consumption Survey 1987-1988 dataset [9]. A list of products that accounted for at least 90% of the population mean intake of the food groups and nutrients of interest was thus selected. Some additional food items were added to the list due to specific hypotheses (e.g. garlic) or expected changes in food patterns (e.g. low-fat products).

The questionnaire contains color photographs of 2 to 4 differently sized portions of 21 foods. The photographs were taken by a professional food photographer under highly controlled conditions of distance, angle, light and presentation. Subjects could choose one of the amounts shown in a photograph or indicate that they ate less than the smallest or more than the largest amount shown. Criteria for the selection of food items to be photographed for the questionnaire were: no natural unit or household measure applicable and a large variation in portion size. For most other items the consumption frequency was asked in number of specified units (slices, glasses, natural units etc.); for a few foods a standard portion size was assumed.

The questionnaire contains blank spaces for filling in brand names of margarine and cooking fat. The food items included in the questionnaire, the types of additional questions posed about these items, and the way in which portions of each food item were estimated are listed in Appendix 1. On average it took the subjects one hour to fill out the questionnaire.

Validation study

Subjects. The validation study was carried out in the pilot-phase of the EPIC study before the actual enrolment of cohort members. The subjects were recruited from two study populations of ongoing projects, in which the Dutch part of the EPIC study was later integrated. The age and sex-stratified sample consisted of 260 women (age 50-70 years) who were invited to take part in a breast cancer screening program in Utrecht and a population of 700 men and women (age 20-60 years) who participated in the Monitoring Program for Cardiovascular Risk Factors in the towns Amsterdam, Doetinchem and Maastricht. For logistical and statistical reasons our aim was to end up with 120 subjects after non-response, subject selection, and drop-out during the study. Of the 960 people invited by mail, 240 (25%) responded positively, 288 (30%) refused to participate, and 432 (45%) did not respond. The main reasons for refusal were lack of time, unable to combine with job, and (among the females in Utrecht) health problems. Out of the 240 subjects that responded positively, we selected 134 subjects equally distributed across the four towns, in 20-year age groups and of both sexes. For 121 subjects, 63 men and 58 women, complete dietary data were obtained; the results presented here pertain to these subjects only. Some characteristics of the subjects are given in table 1. All subjects signed an informed consent form.

Table 1. Description of the subjects of the study population

Characteristics	Males (n=63)		Females (n=58)	
	Mean	sd	Mean	sd
Age (year)	42.6	11.1	49.0	14.6
Height (m)	1.79	0.06	1.65	0.07
Weight (kg)	81.6	9.7	67.9	9.2
Body mass index (kg/m2)	25.5	2.9	24.9	3.5
Change in weight over 13 months (kg)	0.2[a]	2.1	0.5[a]	2.7
Education (frequency %):				
lower vocational & primary	40		57	
intermediate vocational & secondary	32		15	
higher vocational & university	28		28	

[a] No statistically significant difference from 0 (two-sided paired t-test; p>0.1)

Study design and data collection. Data collection started in October 1991 and took 13 months. In months 1, 7 and 13 the food frequency questionnaire was administered in order to test reproducibility. The questionnaire was mailed to the subject and filled out at home. It was checked for completeness and consistency in a standardized way during a visit to the research center.

Relative validity was assessed by comparing the data collected from the questionnaire with that drawn from 24-h dietary recalls repeated 12 times. The 24-h recall interviews were performed monthly during the months 2 through 13 by eight dieticians and nutritionists, according to a standardized protocol; the eight interviewers first underwent thorough training. A summary of the foods recalled and a checklist of items, such as snacks and alcoholic beverages, were part of the interview. The first and seventh recall interviews were conducted at the home of the subject in order to measure the volume of commonly used household tableware and weigh the constituents of one or more sandwiches prepared by the subject. The data obtained were also used for portion size estimates during other recalls. Six 24-h dietary recalls were administered face-to-face and six by telephone. The interviews by telephone took place without prior warning; for the face-to-face interviews this was not feasible. For most of the subjects, the interview days were evenly divided over Mondays to Saturdays. Half of the interviews on Monday concerned the previous Sunday and the other half the previous Saturday. The interviews on Tuesday to Saturday were all about the day before. With a few exceptions due to practical circumstances each subject was interviewed by the same interviewer throughout the study.

Processing the data. For all main food items in the food frequency questionnaire, frequencies per day were calculated first. If the sum of the frequencies for the individual cooked vegetables was not equal to the answer to the question on total cooked vegetable consumption, the frequencies for the individual vegetables were corrected proportionally, as suggested by Haraldsdóttir [7]. This was also done for meat. To convert relative frequencies into absolute frequencies we defined 'always/mostly' as 90% of the absolute frequency of the food item referred to. The categories often, sometimes, and seldom/never were defined as 65, 35 and 10%, respectively. If the sum of these calculations for a set of relative questions was not equal to the frequency of the food item referred to, a proportional correction was made. For example, if a subject reported that the method of preparation of eggs was

always/mostly (=90%) boiled and was sometimes (=35%) fried, this would be corrected to 72% and 28% of the reported number of eggs. In this way the consumption frequencies for 178 foods were estimated. When the answer 'smaller than the smallest portion' was chosen for a photographic question, the portion was assumed to be 50% of the smallest portion; for 'larger than the largest portion' it was 125% of the largest portion. Frequencies per day and portion sizes were multiplied to obtain grams per day for each food item. Then food items were grouped into 20 food groups, i.e. 16 main food groups and two subgroups for vegetables (raw/cooked) as well as fruits (citrus/non-citrus).

The 24-h dietary recalls were coded by the interviewers, according to Dutch national coding instructions [10]. These foods were also grouped into 20 food groups. Weighted means of the 12 24-h food group intakes were calculated, with a weight of two for Saturdays and Sundays and a weight of one for the other days. In this way a correction was made for the underrepresentation of Saturdays and Sundays.

Statistical analyses. Statistical analyses were performed with SAS-software version 6.07 for men and women separately.

Since food group intakes were generally skewed towards higher values, we chose to use non-parametric statistics. To compare absolute group intakes, sample medians and 25th and 75th percentiles were computed. Relative validity at the group level was assessed by examining differences in distributions of intake between the food frequency questionnaires and the 24-h recalls. The sign test was used to test whether the differences in distributions were statistically significant, which was defined as two-sided p-values ≤ 0.05. Results are presented for the questionnaire at the start of the study as for this questionnaire the participation in the study could not have influenced the answers.

The reproducibility of the ranking of subjects according to food group intake was expressed as Spearman rank correlation coefficients between food group intakes based on the first questionnaire and those based on the second and third questionnaires. Spearman correlation coefficients between the intake estimates based on the questionnaire and those based on the 24-h recalls were used as a measure of the relative validity of the ranking of subjects. The food groups fish, eggs, nuts & seeds, raw vegetables, citrus fruit, and alcoholic beverages were not reported in many single recalls and 12 repeated 24-h recalls therefore represent an inaccurate estimation of

habitual consumption at the individual level. Therefore, for these food groups we also determined tertiles of intake, as defined by the food frequency questionnaire, and ascertained whether the mean recalled intake increased per tertile.

The influence of the food photographs in the questionnaire on the relative validity of the food group intakes was evaluated as follows: food group estimates were recalculated by substituting the size of the individual portion obtained from the questions concerning food photographs for a standard portion size. This standard was the amount shown in the middle photograph or, in case of an even number of photographs, the average of the middle two. The distributions of food group intakes thus calculated were compared with those calculated originally and with those based on the 24-h dietary recalls, using the sign test. Furthermore Spearman correlation coefficients between the food group estimates calculated in both ways and those based on the 24-h dietary recalls were compared.

Results

Characteristics of the study population are shown in table 1. The 121 subjects who provided complete dietary information did not significantly change in weight during the 13-month study period. Average body mass index for males was 25.5 kg/m^2, for females 24.9 kg/m^2.

Absolute intake on the group level

Daily median intakes for 8 food groups assessed by the 24-h recalls and by the first food frequency questionnaire are given in table 2. For these food groups the questionnaire estimates were only based on reported frequencies and not on reported portion sizes. For men, the median estimate of alcoholic beverages according to the questionnaire was 66% lower than that based on the 24-h recalls, which means an underestimation of almost one bottle of beer per day. For women the questionnaire yielded a 14% higher estimate of median bread intake.

Table 2. *Daily median (P25, P75) food group intake (g) estimated by means of 12 24-h dietary recalls and the Dutch EPIC food frequency questionnaire (FFQ)*

Food group	Males (n=63)		Females (n=58)	
	24-h recall	FFQ	24-h recall	FFQ
Bread	168 (122,198)	173 (115,240)	97* (80,123)	111 (93,133)
Fruit	124 (71,205)	129 (81,198)	138 (90,210)	157 (92,258)
citrus	18 (1,66)	27 (14,51)	26 (13,54)	42 (21,80)
non-citrus	103 (57,160)	100 (60,171)	114 (65,154)	112 (67,182)
Fish	6 (0,25)	8 (4,14)	6 (0,15)	8 (2,14)
Eggs	13 (8,20)	14 (8,21)	12 (6,17)	14 (7,18)
Non-alc. beverages	1341 (1106,1814)	1470 (1088,2010)	1285 (1048,1462)	1305 (1099,1535)
Alcoholic beverages	419* (95,735)	143 (57,414)	59 (5,163)	52 (6,128)

* $p \leq 0.05$ for two-sided sign test comparing intake based on 12 24-h recalls with that based on the food frequency questionnaire at the start of the study.

The daily median intake estimates for 12 food groups are shown in table 3. These food group intakes assessed by the food frequency questionnaire were based on reported frequencies and the amounts indicated by referring to portion sizes shown on photographs. Median estimates of milk & milk products and added fats for both sexes and cereals & pasta for women obtained from the food frequency questionnaire were higher than those based on the 24-h recalls. The opposite was observed for biscuits & pastry for both men and women and for potatoes for women.

Replacing individual portion size information in the questionnaire data by standard portion sizes changed 10 out of the 12 distributions of food group estimates significantly for men (table 3). Median intake estimates for 7 of the 10 food groups became lower, whereas those for raw vegetables, cheese, and added fats became higher. For women, the distributions of 9 food group estimates changed significantly when portion size information was replaced by standard portion sizes. In contrast to males, most of the nine median intake estimates became higher, with the exceptions of milk & milk products and sugar & sweet products. For both men and women, most food group estimates based on standard portion sizes deviated more from the 24-h recall estimates than those based on food photograph portions. Milk & milk products was an exception to this.

Table 3. *Daily median (P25,P75) food group intake (g) estimated by means of 12 24-hr dietary recalls, the Dutch EPIC food frequency questionnaire (FFQ photo) and the same questionnaire after replacing answers related to the food photographs by standard portion sizes (FFQ standard)*

food group	Males (n=63)						Females (n=58)					
	24-hr recalls		FFQ photo		FFQ standard		24-hr recalls		FFQ photo		FFQ standard	
Cereals & pasta	37	(16,72)	55[b]	(25,86)	48	(26,68)	31[ab]	(18,49)	34[b]	(20,59)	42	(28,64)
Potatoes	131[b]	(86,196)	147	(79,188)	161	(136,190)	82[ab]	(58,133)	66[b]	(41,102)	149	(100,169)
Vegetables	156	(134,223)	155	(119,214)	153	(127,205)	132[b]	(100,172)	152[b]	(109,201)	170	(135,226)
raw vegetables	40[b]	(22,65)	43[b]	(22,70)	54	(38,86)	34[b]	(17,54)	35[b]	(18,60)	56	(29,94)
cooked vegetables	107	(86,167)	109[b]	(84,150)	106	(82,128)	89[b]	(67,133)	105[b]	(78,140)	119	(89,145)
Meat	138[b]	(109,180)	131[b]	(107,160)	117	(99,134)	93[a]	(71,128)	85	(52,124)	115	(100,162)
Cheese	30[b]	(20,49)	35[b]	(17,58)	49	(19,70)	29[b]	(20,42)	29[b]	(17,43)	45	(26,63)
Milk & milk products	272[a]	(138,535)	370[b]	(238,596)	348	(202,591)	312[ab]	(166,386)	378[b]	(250,516)	349	(228,516)
Added fats	32[ab]	(20,39)	47[b]	(28,62)	57	(35,73)	20[ab]	(13,30)	35[b]	(22,42)	49	(33,58)
Sugar and sweet products	57	(26,89)	50[b]	(25,86)	46	(25,82)	30[b]	(21,50)	24[b]	(12,47)	22	(11,42)
Biscuits & pastry	48[ab]	(25,73)	32[b]	(23,61)	30	(22,49)	50[ab]	(27,69)	41	(19,55)	38	(19,55)
Nuts & seeds	9	(3,17)	11[b]	(5,22)	9	(4,15)	4	(1,10)	6	(2,18)	6	(2,14)

[a] = significantly different from food frequency data using food photographs for portion information (p-value sign test ≤0.05)
[b] = significantly different from food frequency data using standard portion sizes (p-value sign test ≤0.05)

Ranking of subjects

The reproducibility and relative validity for the ranking of subjects, expressed as Spearman correlation coefficients, are presented in table 4. For men, the median 6-month reproducibility was 0.76, ranging from 0.49 for fish to 0.91 for alcoholic beverages; the median 12-month reproducibility was 0.71. For women, median reproducibility was 0.76 at 6 months (range 0.61-0.91) and 0.77 at 12-months. If non-drinkers were excluded, the correlation coefficients for alcoholic beverages were slightly lower than those in table 4 (0.86 and 0.77 for men, 0.84 and 0.86 for women).

Table 4. *Spearman rank correlation coefficients between food group estimates based on repeated food frequency questionnaires (reproducibility) and between estimates based on the food frequency questionnaire and 12 24-h recalls (relative validity)*[a]

Food group	Males (n=63)				Females (n=58)			
	reproducibility		validity		reproducibility		validity	
	6-mth	12-mth	ffq[b]	standard[c]	6-mth	12-mth	ffq	standard
Bread	0.90	0.86	0.76	-	0.88	0.85	0.78	-
Cereals & pasta	0.81	0.79	0.51	0.48	0.83	0.80	0.67	0.67
Potatoes	0.86	0.85	0.58	0.45	0.78	0.75	0.70	0.55
Vegetables	0.80	0.76	0.38	0.36	0.61	0.65	0.31	0.25
raw vegetables	0.78	0.67	0.49	0.41	0.61	0.68	0.32	0.39
cooked vegetables	0.71	0.69	0.21	0.23	0.68	0.67	0.41	0.38
Fruit	0.70	0.61	0.68	-	0.77	0.77	0.56	-
citrus fruit	0.70	0.68	0.53	-	0.75	0.77	0.41	-
non-citrus fruit	0.72	0.61	0.72	-	0.75	0.77	0.62	-
Meat	0.71	0.68	0.47	0.39	0.77	0.80	0.70	0.59
Fish	0.49	0.45	0.32	-	0.61	0.63	0.37	-
Eggs	0.71	0.72	0.41	-	0.82	0.79	0.43	-
Cheese	0.77	0.71	0.64	0.56	0.67	0.70	0.38	0.32
Milk & milk products	0.85	0.73	0.71	0.69	0.75	0.78	0.79	0.77
Added fats	0.74	0.76	0.65	0.65	0.80	0.76	0.67	0.53
Sugar and sweet products	0.87	0.71	0.78	0.76	0.80	0.79	0.69	0.69
Biscuits & pastry	0.56	0.67	0.56	0.52	0.74	0.76	0.45	0.50
Nuts & seeds	0.73	0.70	0.65	0.59	0.80	0.73	0.38	0.35
Non-alcoholic beverages	0.77	0.69	0.67	-	0.62	0.74	0.49	-
Alcoholic beverages	0.91	0.83	0.74	-	0.91	0.92	0.87	-

[a] 95% confidence intervals with n=60 are -0.06-0.43 for r=0.2; 0.05-0.51 for r=0.3; 0.16-0.59 for r=0.4; 0.28-0.67 for r=0.5; 0.41-0.74 for r=0.6; 0.54-0.81 for r=0.7; 0.69-0.88 for r=0.8; 0.84-0.94 for r=0.9; [b] data from the Dutch EPIC questionnaire as it is; [c] data from the Dutch EPIC questionnaire after substituting photograph portion size information for standard portion sizes

Median Spearman correlation coefficients between the food group estimates based on the first food frequency questionnaire and those based on the 24-h recalls were 0.61 for men and 0.53 for women. For the second and third food frequency questionnaires the correlation coefficients were similar to those reported in table 4 with medians of 0.64 and 0.60 for men and medians of 0.58 and 0.52 for women. For the first food frequency questionnaire the correlation coefficients ranged from 0.21 for cooked vegetables to 0.78 for sugar & sweet products for men. For women, the lowest correlation coefficient of 0.31 was observed for vegetables. The highest correlation coefficient of 0.87 for alcoholic beverages for women decreased to 0.76 when non-drinkers were excluded (n=54), whereas excluding non-drinkers did not change the correlation coefficient for men (n=45).

Table 5. *Mean daily intake (g) of food groups not generally eaten on a daily basis based on the Dutch EPIC food frequency questionnaire (ffq) and 12 24-h dietary recalls per tertile of intake (T) as defined by the food frequency questionnaire.*

		Men (n=63)			Women (n=58)		
Food group	Method	T1	T2	T3	T1	T2	T3
Citrus fruit	ffq	9.4	26.9	68.5	14.1	44.4	110.5
	24-h recalls	9.9	27.4	66.3	18.4	42.5	49.8
Raw vegetables	ffq	17.2	44.2	109.8	12.7	35.0	82.8
	24-h recalls	24.5	50.2	71.1	30.6	39.1	46.5
Fish	ffq	2.3	7.1	18.4	1.0	8.0	17.2
	24-h recalls	8.6	15.4	24.6	5.1	10.8	15.1
Eggs	ffq	5.7	13.0	33.0	5.2	12.7	21.2
	24-h recalls	11.4	18.8	26.2	8.7	18.6	18.0
Nuts & seeds	ffq	3.9	12.5	32.1	1.7	6.5	25.8
	24-h recalls	4.0	14.0	25.1	4.1	5.3	11.3
Alcoholic	ffq	24.6	165.3	744.7	2.7	50.9	226.4
beverages	24-h recalls	173.7	475.9	1013.5	7.7	96.6	226.6

For six food groups which are not generally eaten on a daily basis, the mean intake estimates based on the 24-h recalls and the dietary questionnaire are presented for subgroups as defined by tertiles of intake assessed from the questionnaire; see table 5. With the exception of egg intake for women, the mean intake based on 24-h recalls increased per tertile. The differences between tertiles were in general smaller for mean

intakes based on 24-h recalls than for those based on the food frequency questionnaire, although this was not the case for alcoholic beverages for men.

Table 4 also shows the influence of the photographs in the food frequency questionnaire on the relative validity for ranking subjects according to food group intake. When the portion sizes assessed by food photographs were replaced by standard portion sizes, Spearman correlation coefficients between food group estimates based on the food frequency questionnaire and those based on 24-h dietary recalls decreased by more than 0.10 for potatoes for both men and women and for meat and added fats for women. On the other hand, increases of 0.05 or more were observed for raw vegetables and biscuits & pastry for women. The median result was a decline of 0.03 in correlation coefficients.

Discussion

In the present study, the reproducibility and relative validity for habitual food group intakes estimated by a Dutch food frequency questionnaire were explored. Since the study was a pilot phase of the EPIC study, we recruited volunteers from ongoing-projects on which the Dutch EPIC study was later grafted. Requirements for participation were considerably high, as the study lasted 13 months and comprised of monthly interviews, the repeated filling out of extensive questionnaires, and quarterly collection of 24-h urine and blood samples. This probably explains the rather low positive response of 25% to our invitation to participate. It could well be that the ability of these selective subjects to describe their food habits was somewhat better than that of average cohort participants [11]. Furthermore, the limited number of subjects (63 men and 58 women) resulted in confidence intervals for the correlation coefficients which were rather broad: e.g., a Spearman correlation coefficient of 0.6 has a 95% confidence interval of 0.41-0.74 (n=60).

The reproducibility, when expressed as Spearman correlation coefficients, was generally good. Median correlation coefficients for both men and women and the 6 or 12-month reproducibility exceeded 0.7. Poor reproducibility ($r<0.50$) was only observed for ranking men according to fish intake. This was possibly due to the large proportion of men who consumed fish less then once per week. Because of this, small absolute differences in the reported consumption frequency of fish between the questionnaires could have caused larger differences in the ranking. Consequently, a low relative validity was also found for this food group for men ($r=0.32$).

Two other Dutch studies on the reproducibility of food frequency questionnaires yielded similar median correlation coefficients of about 0.7 for food groups [12,13]. The reproducibility of individual food intake estimated by means of food frequency questionnaires is generally more variable with a higher percentage of poor correlations than our results for aggregated food groups [14-16].

To assess relative validity, we chose repeated 24-h dietary recalls as reference method. In contrast to the food frequency questionnaire this method involves an open interview and its use does not depend on the long-term memory of subjects and their ability to average food intake over a longer period. Furthermore, to estimate portion sizes other methods were used than those in the questionnaire. Hence, it seems reasonable to assume that there is little correlation between measurement errors in estimated food group intakes based on the food frequency questionnaire and those assessed by the 24-h recalls. This implies that the correlation coefficients are probably not inflated to a great extent. Only the tendency of some subjects to underreport or overreport, irrespective of method, may have counteracted this. On the other hand, correlation coefficients might be artificially low because of measurement errors in the 24-h recalls and large intraindividual day-to-day variation for some food groups. There was no indication that additional measurement error in the form of underreporting has occurred by recalling Saturday's diet on Mondays, as mean energy intake on Saturdays was highest of all days of the week. Some measurement error might have been caused by interobserver variation as 5 of the 28 pairwise comparisons between mean energy intake differed significantly from each other ($p<0.025$). Deattenuation of correlation coefficients for the day-to-day variation was not possible since most variables were highly skewed and included many zero-values. Instead we evaluated whether the mean intake of those food groups not usually consumed on a daily basis assessed by 24-h recalls increased per tertile of intake, as defined by the food frequency questionnaire. With the exception of egg intake for women this turned out to be the case, which means that the questionnaire can be used to differentiate between groups with low, moderate and high intakes of food groups.

The observed median Spearman correlation coefficients of 0.61 for men and 0.53 for women between food group intakes based on the questionnaire and those based on the reference method are comparable to those described for other food frequency questionnaires [11,13,17,18]. Precise comparisons are difficult, however, because of methodological differences, differences in populations, and differences in

food groups reported. Although Spearman correlation coefficients have their limitations for the interpretation of relative validity [19], comparison of the classification of subjects into tertiles on the basis of the two methods (data not shown) did not lead to other conclusions. The results for the second and third dietary questionnaires were close to those found for the first. This indicates that neither learning effect nor synchronization of the periods of reference for the 24-h recalls and the food frequency questionnaire was an issue in our population. In table 6 our results on relative validity for vegetables and fruits, which are of prime interest for the EPIC study, are compared with those of five other studies. It is clear that, like in our study, the relative validity for the ranking of individuals according to vegetable intake is generally poor. The highest correlation coefficients observed by Bloemberg *et al* [13] were probably overestimated because they used a dietary history as reference method which is based on the same concept as the food frequency questionnaire. A small interindividual variation in consumption frequency and measurement errors in portion size estimation might be possible reasons for the generally low relative validity for estimates of vegetable intake. For fruit the correlation coefficients are generally reasonably good, although values below 0.4 were reported by Hankin *et al.* [20] and Nes *et al.* [18].

Table 6. *Overview of the reported relative validity for ranking individuals according to vegetable and fruit intake assessed by self-administered food frequency questionnaires*

Reference	Vegetables		Fruit		Type of correlation	Reference method
	men	women	men	women		
Pietinen *et al.* 1988	0.53	-	0.66	-	Pearson	12 x 2-day record
Hankin *et al.* 1991	0.39	0.19	0.60	0.34	Intraclass	4 x 1-week record
Nes *et al.* 1992	-	0.42	-	0.38	Spearman	14-day record
Bloemberg *et al.* 1993	0.51	0.63	0.73	0.64	Spearman	dietary history
Goldbohm *et al.* 1994	--0.38--[a]		--0.60--		Spearman	3 x 3-day record
This study	0.38	0.31	0.68	0.56	Spearman	12 x 24-h recall

[a] men and women together

In order to interpret the results on the relative validity of food group estimates at the group level, the possibility of a systematic bias in the reference method should be considered. In the literature acceptable group means obtained from standardized 24-h dietary recalls have been reported by some, but underestimates compared to diet

records have also been observed [21]. Comparison of our mean intakes with those found with the Dutch National Food Consumption Survey 1987/1988 based on 2-day records [22] revealed similar values for males, but for females we found 5 to 10% lower means for most of the food groups.

In our study, higher as well as lower median food group intakes as estimated by the food frequency questionnaire were observed in comparison with the 24-h dietary recalls. Statistically significant higher median intakes for both men and women were found for milk & milk products and added fats and lower medians for biscuits & pastry. For males the median for alcoholic beverages was only 34% of the recall estimate. This means an underestimation of almost one bottle of beer per day. Increasing the size of a standard beer glass from 200 to 225 g learned us that the underestimation was mainly caused by an underestimation of the number of glasses usually consumed, because underestimation remained large. Other validity studies have also shown both overestimates and underestimates for food frequency questionnaires [11,13,17]. As it cannot be expected that dietary data obtained by means of different questionnaires among the cohorts of the EPIC study will exhibit similar bias, the need for a calibration study to correct for these effects in intercohort analyses is obvious [5].

Food photographs

The need for information on portion sizes in food frequency questionnaires has been debated. One could argue that fixed portion sizes 'over-standardize' consumption estimates. On the other hand it has been shown that in contrast to frequency, portion size has a narrow range of variation [23] and that intraindividual variation in portion size is high compared to interindividual variation [2].

In our questionnaire we incorporated food photographs to estimate habitual portion sizes for those foods that could not easily be assessed in natural units or household measures and which showed a high variation in portion size. Analysis of our data without the answers related to the food photographs, using standard portion sizes instead, generally resulted in average food group intakes that deviated more from those of the 24-h dietary recalls. For males, the intakes generally decreased versus an increase for females. For milk & milk products the photograph of dairy desserts caused an overestimation of intake by both men and women. Whether the actual amounts shown in the photographs were perceived as smaller or subjects overestimated the amount they usually consumed is not clear. Our findings are in accordance with the

fact that men usually consume larger portions than women [24]. Tjønneland *et al.* [3] who reported the effects of substituting answers based on photographic portions for standard portion sizes for five food groups did not find mean intakes that consistently deviated more from the reference method.

In this study the effects of replacing answers based on photographic portions by standard portion sizes on the relative validity for the ranking of individuals were usually but not always in a negative direction and were generally small. A decrease of more than 0.10 in the Spearman correlation coefficient was observed for potatoes for both sexes and for meat and added fats for women. Hankin *et al.* [25] observed a similar effect for 30 food items. They found a shift in the average correlation coefficient from 0.59 to 0.55. In the study by Tjønneland *et al.* [3] the correlation coefficients decreased for 4 out of 5 food groups among men and for 3 food groups among women when standard portion sizes were introduced. The largest decrease was found for fish (0.07) among men and for potatoes (0.09) among women. Other published studies on the effect of information about portion size in food frequency questionnaires on the relative validity for the ranking of subjects according to nutrient intake generally have shown small effects [3]. Whether post-hoc evaluations of the effect of food photographs will give the same results as comparing a questionnaire with and one without photographs is uncertain.

Conclusion

In conclusion, the EPIC food frequency questionnaire seems reasonably valid for ranking individuals according to food group intake, although considerable measurement error was observed for some food groups, such as fish and vegetables. The food photographs in the questionnaire contributed little to the relative validity of the ranking. Population level estimates of food group intake varied in relative validity. For both sexes, intake of milk & milk products and added fats were overestimated by the questionnaire, whereas biscuits & pastry intake was underestimated. For men, alcohol intake was seriously underestimated. The food photographs generally had a positive influence on the relative validity for absolute food group intake. Given these results we feel confident in using a version of this questionnaire for the Dutch EPIC cohorts to rank subjects according to food group intake.

Acknowledgements

The authors thank all subjects in the study for their time and effort. We are grateful to Gatske Obermann-de Boer, Petra Peeters and Jet Smit for their continuous support during the planning and data collection phase. Karin Blank, Anneke Blokstra, Sigrid Braak, Anja Muusers and Lucie Viet are acknowledged for the dietary interviewing. This study was part of the EPIC project and was supported financially by the Europe Against Cancer Program of the European Union and by the Ministry of Welfare, Health and Culture in the Netherlands.

References

1. Willett W. Nutritional Epidemiology, Monographs in Epidemiology and Biostatistics. Vol. 15. New York, Oxford: Oxford University Press, 1990
2. Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, Willett WC. Variability in portion sizes of commonly consumed foods among a population of women in the United States. Am J Epidemiol 1988; 127: 1240-9
3. Tjønneland A, Haraldsdóttir J, Overvad K, Stripp C, Ewertz M, Jensen OM. Influence of individually estimated portion size data on the validity of a semiquantitative food frequency questionnaire. Int J Epidemiol 1992; 21: 770-7
4. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol 1992; 3: 783-91
5. Kaaks R, Plummer M, Riboli E, Estève J, Van Staveren W. Adjustment for bias due to errors in exposure assessment in multicenter cohort studies on diet and cancer: a calibration approach. Am J Clin Nutr 1994; 59: (suppl), 245S-50S
6. Ocké 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. (submitted)
7. Haraldsdóttir J. Minimizing error in the field: quality control in dietary surveys. Eur J Clin Nutr 1993; 47: (suppl 2), S19-S24
8. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986; 124: 453-69
9. Hulshof KFAM, Van Staveren WA. The Dutch National Food Consumption Survey: design, methods and first results. Food Policy 1991; 16: 257-60
10. Hulshof KFAM, Van der Heijden LJM, Donders-Engelen M. Measures, weights, and code numbers 1992. V92.003. Zeist: TNO-Voeding and Landbouwuniversiteit Wageningen, 1992. (In Dutch)
11. Goldbohm RA, Van den Brandt PA, Brants HAM, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr 1994; 48: 253-65
12. Bueno de Mesquita HB, Smeets FWM, Runia S, Hulshof KFAM. The reproducibility of a food frequency questionnaire among controls participating in a case-control study on cancer. Nutr Cancer 1992; 18: 143-56

13. Bloemberg BPM, Kromhout D, Jansen AM, Goddijn HE. Reproducibility and validity of a short self-administered semi-quantitative food frequency questionnaire. In: Bloemberg B P M. On the effect of measurement error in nutritional epidemiology using dietary history and food frequency methodology. Thesis. Leiden: Leiden University, 1993, pp. 45-65
14. Nomura A, Hankin JH, Rhoads G. The reproducibility of dietary intake data in a prospective study of gastrointestinal cancer. Am J Clin Nutr 1976; 29: 1432-6
15. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. Am J Epidemiol 1988; 128: 667-76
16. Salvini S, Hunter DJ, Sampson L, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. Int J Epidemiol 1989; 18: 858-67
17. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. Am J Epidemiol 1988; 128: 655-66
18. Nes M, Frost Andersen L, Solvoll K, et al. Accuracy of a quantitative food frequency questionnaire applied in elderly Norwegian women. Eur J Clin Nutr 1992; 46: 809-21
19. Bellach B. Remarks on the use of Pearson's correlation coefficient and other association measures in assessing validity and reliability of dietary assessment methods. Eur J Clin Nutr 1993; 47 (suppl 2): S42-5
20. Hankin JH, Wilkens LR, Kolonel LN, Yoshizawa CN. Validation of a quantitative diet history method in Hawaii. Am J Epidemiol 1991; 133: 616-28
21. Kuskowska-Wolk A. Recall methods - a review of the methodology studies. Vår Föda 1986; 4 (suppl): 219-49
22. Ministerie van Welzijn, Volksgezondheid en Cultuur & Ministerie van Landbouw en Visserij What do the Dutch eat. Results from the National Food Consumption Survey 1987-1988 (In Dutch). Rijswijk, 1988
23. Samet JM, Humble CG, Skipper BE. Alternatives in the collection and analysis of food frequency interview data. Am J Epidemiol 1984; 120: 572-81
24. Caster WO. Systematic estimation of food intakes from food frequency data. Nutr Res 1986; 6: 469-72
25. Hankin JH, Rhoads G, Glober GA. A dietary method for an epidemiologic study of gastrointestinal cancer. Am J Clin Nutr 1975; 28: 1055-61

Appendix 1. *The Dutch EPIC food frequency questionnaire: food items, additional questions, and type of portion size estimation*

Food item	Additional question	Portion size estimation
Breakfast cereals	addition of dairy products	number of spoons, standard proportion for dairy product
White bread	-	number of slices
Whole wheat bread	-	number of slices
Wheat bread	-	number of slices
Currant/raisin bread	-	number of slices
Rye bread	-	number of slices
Buns, rolls, croissant	-	number of slices
Dutch rusk, crackers	-	number of slices
(for total of bread)	types of margarine/butter	photographs
Cheese on bread	subtypes	number of slices & photographs
Cold cuts on bread	subtypes	number of slices & photographs
Egg on bread	preparation	number of natural units
Sweet sandwich spreads	subtypes	number of slices & photographs
Coffee	subtypes	number of cups
	additions	number of lumps for sugar; photographs for milk
Tea	additions	number of cups for tea, lumps for sugar, standard proportion for milk
Buttermilk	-	number of glasses
Drinking yoghurt	-	number of glasses
Milk	subtypes	number of glasses
Soup	-	number of plates/cups
Rice	subtypes	photographs
Pasta	-	photographs
Hot sauces	subtypes	number of spoons
Boiled potatoes	-	photographs
French fries during meal	preparation	photographs
French fries beside meal	-	standard portion
Roasted potatoes	-	more/equal/less than potatoes
Mayonnaise etc.	subtypes	number of spoons
Garlic	-	-
Raw vegetables[a]	subtypes	photographs
	additions	standard proportion
Roasted vegetables	subtypes	photographs
Total boiled vegetables[a]	additions	standard proportion for additions
Boiled string beans/broad beans	-	photographs
Peas	-	photographs
Other legumes	-	photographs
Boiled red beets	-	photographs
Boiled cabbage	-	photographs
Boiled spinach	-	photographs
Boiled endive	-	photographs

Boiled leek or onion	-	photographs
Boiled carrots	-	photographs
Apple compote	-	number of spoons
Total meat (other than cold cuts)	-	photographs
Ground meat	subtypes	same as total meat
Beef	subtypes	same as total meat
Pork	subtypes	same as total meat
Organ meats	-	same as total meat
Smoked sausage	-	same as total meat
Chicken	subtypes	same as total meat
Other meat	-	same as total meat
Gravy	preparation	photographs
Eggs at dinner	preparation	number of natural units
Fish, molluscs, shrimps	subtypes	standard portion
	preparation	
Meals without meat/fish/egg	types of replacers	standard portion
(for all frying/roasting:)	types of cooking fat	standard proportion
Dairy desserts	subtypes	photographs
	additions	standard portion
Fruit[a]	subtypes	number of natural units
Fruit & vegetable juice	subtypes	number of glasses
Tap water	-	number of glasses
Non-alcoholic beverages	subtypes	number of glasses
Alcohol free beer	-	number of glasses
Beer	-	number of glasses
White wine	-	number of glasses
Red wine, rosé wine	-	number of glasses
Port, sherry, vermouth, advocaat	-	number of glasses
Spirits	-	number of glasses
Chocolates, bonbons	-	number of units
Chocolate bars, candy bars	-	number of units
Liquorice	-	number of units
Other sweets, toffees, acid drop	-	number of units
Honey bread	butter/margarine	number of units, standard portions for butter/margarine
Apple pie, fruit pies	-	number of units
Whipped cream cake	-	number of units
Cake, large cookies	-	number of units
Biscuits, small cookies	-	number of units
Russian salad	-	number of units
Meat snacks	-	number of units
Spring rolls	-	number of units
Peanuts & other nuts	-	photograph
Cheese, as snack	-	standard portion
Sausage, as snack	-	standard portion
Salty snacks	-	photographs

[a] separate questions for summer and winter

Chapter 6

The Dutch EPIC food frequency questionnaire II Relative validity and reproducibility for nutrients

Marga Ocké, Bas Bueno de Mesquita, Margreet Pols, Henriette Smit, Wija van Staveren, Daan Kromhout.

Int J Epidemiol, in press

Abstract

A self-administered semi-quantitative food frequency questionnaire was developed for the Dutch cohort of the EPIC study. The reproducibility and relative validity of nutrient intake as assessed by this questionnaire were investigated in a population of 121 men and women. To assess the relative validity 12 monthly 24-h recalls served as reference method, together with four determinations of 24-h urinary nitrogen excretion, predicted basal metabolic rate, and serum ß-carotene and α-tocopherol levels. Protein and, among women, energy intake were underestimated by the questionnaire compared to urinary nitrogen excretion and the basal metabolic rate, respectively. The underestimation for protein decreased with increasing protein intake. Pearson correlation coefficients between nutrient intakes assessed by repeated questionnaires ranged from 0.70 to 0.94 among men and from 0.59-0.94 among women. Correlation coefficients between nutrient intakes assessed by the questionnaire and 24-h recalls ranged from 0.26-0.83 for men and 0.35-0.90 for women, with medians of 0.59 and 0.58, respectively. Correlation coefficients between 0.2 and 0.5 were observed for ß-carotene and vitamin C for men and for ß-carotene and vitamin E for women. Associations with serum ß-carotene (r=-0.16 for men; 0.13 for women) and α-tocopherol (0.23 and 0.15, respectively) were much poorer than those obtained with 24-h recalls. Correlations between protein intake and 24-h urinary nitrogen excretion were 0.47 and 0.49, respectively. In conclusion, the food frequency questionnaire seems adequate for ranking subjects according to intake of energy, macronutrients, dietary fibre and retinol, but it does not yield such good results for ß-carotene, vitamin C for men, and vitamin E for women.

Introduction

Food frequency questionnaires are often used to assess habitual dietary intake in epidemiological studies on diet and chronic diseases. This is also the case for the Dutch component of the European Prospective Investigation into Cancer and Nutrition (EPIC), a multicohort study in seven European countries [1]. Since the Dutch EPIC questionnaire was newly developed, it needed to be validated before use in the cohort population. This need is emphasized by the diverse results of validation studies of other food frequency questionnaires [2,3].

The Dutch EPIC food frequency questionnaire is described in a companion paper in this issue, and its relative validity for food groups as assessed in a pilot study among 121 men and women in 1991/1992 is reported [4]. In the present paper, assessment of the reproducibility and relative validity for energy, macronutrients, dietary fibre, retinol, ß-carotene, vitamin C, and vitamin E, as determined in the same pilot study, is described. Due to the lack of a true "gold standard", we used multiple reference methods to measure relative validity, i.e. 12 monthly 24-h recalls, four determinations of 24-h urinary nitrogen excretion, predicted basal metabolic rate, and ß-carotene and α-tocopherol concentrations in serum.

Material and methods

Study design and data collection

The subjects for the validation study were recruited from study populations of ongoing projects in four towns in the Netherlands [4]. The men were 20-60 years of age; the women 20-70 years of age. Of the 960 individuals invited to participate 240 (25%) responded positively. Out of these, we selected 134 subjects, about equally distributed over the four towns, in 20-year age groups, and between both sexes. The results reported in this paper pertain to 121 subjects, 63 men and 58 women, who provided complete dietary data.

An extensive description of the questionnaire is given in the companion paper [4]. In short, the food frequency questionnaire, which is self-administered, contains questions on the habitual consumption frequency during the past year for 79 main food items. Answers can be given in times per day, per week, per month, or per year. Additional questions are asked about the consumption frequency for different subitems, preparation methods, or additions. For 28 food items, questions refer to portion sizes shown in color photographs. For most other items the portion size was specified in the

questionnaire. The questionnaire contains blank spaces for filling in brand names of margarine and cooking fat. In total, the average daily consumption of 178 foods is estimated by means of the questionnaire.

Data collection started in October 1991 and lasted 13 months. The dietary assessment methods used during the study are presented in figure 1. In order to assess reproducibility, the food frequency questionnaire was administered three times; at the start of the study and 6 and 12 months later. The relative validity of the questionnaire was evaluated at two levels: the ability to assess the absolute intake at the group level, and the ability to correctly rank individuals according to nutrient intake. Since the relative validity of the questionnaire when first administered will be most representative of the relative validity for the EPIC cohort, we focused on the relative validity of the first questionnaire.

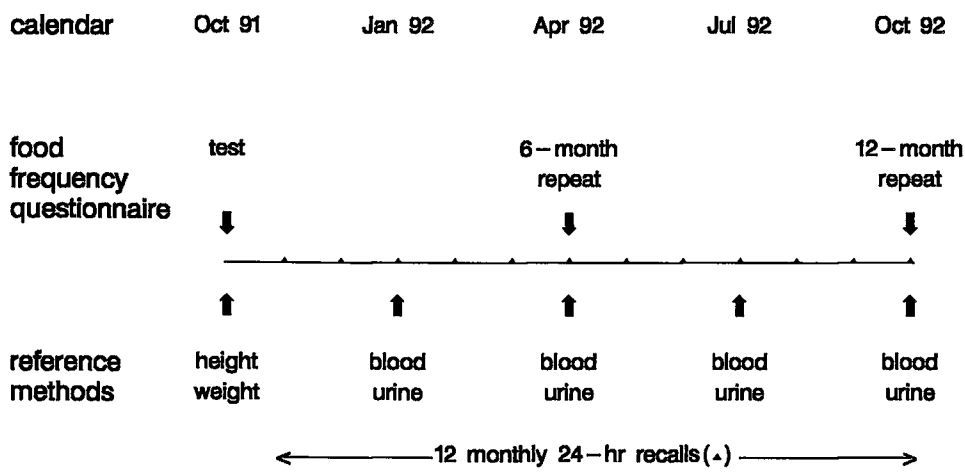


Figure 1. *Time sequence of the validation study*

The main reference method used to assess relative validity consisted of 12 monthly 24-hour dietary recalls. For all but two subjects, the recall days included one Saturday and one Sunday; the other days of the week were on average recalled twice. An extensive description of the 24-h recall interview protocol is given in the companion paper [4].

A second reference method was used to assess the relative validity of protein intake: the repeated 24-h urinary nitrogen excretion. Subjects collected 24-h urine samples four times at three-month intervals. Collection started and ended at the research center in order to monitor the period of urine collection. Afterwards, subjects were asked whether they had lost any specimens of urine during the collection period. Urinary nitrogen concentrations were determined by means of an automated chemical system with a Mitsubishi Total Nitrogen Analyzer TN-05 at the Laboratory for Chemical Analyses of the University Hospital in Leuven (Belgium). The intraindividual coefficient of variation (CV) for blind duplicate samples was 3.0%. The CV's for nitrogen concentrations which were measured daily in Lyphocheck I and II quantitative urine control samples (Biorad Laboratories, Nazareth, Belgium) during the period of analysis were 6.7 and 2.9%, respectively.

The average ratio of energy intake to predicted basal metabolic rate (BMR) was used to estimate relative validity for mean energy intake. A mean value statistically significantly below 1.55 would point to underestimation of energy intake, as even in sedentary living populations this is physiologically unlikely [5]. Height was measured with a wall-mounted stadiometre. The respondents were weighed wearing indoor clothing with empty pockets and without shoes.

Serum α-tocopherol and ß-carotene levels were used to assess the relative validity for the ranking of individuals according to intake of vitamin E and ß-carotene. For this purpose blood specimens from nonfasting subjects were collected four times during the study. The vacutainer tubes were kept for 1 to 4 hours in closed boxes before separating serum. Serum aliquots were stored for up to seven weeks at -20 °C and then transported to a -80 °C freezer. Serum α-tocopherol and ß-carotene concentrations were measured at the Biochemistry Laboratory of the University Hospital in Grenoble (France). A fluorometric technique, adapted from Vatassery & Mortenson [6], was used to measure α-tocopherol; ß-carotene was analyzed by the HPLC technique, adapted from Craft *et al.* [7]. The CV's for quality control samples during this period were 5.6% for α-tocopherol and 12.5% for ß-carotene. Serum total cholesterol was determined enzymatically using a Boehringer testkit [8] at the Clinical Chemical Laboratory of the University Hospital 'Dijkzigt' in Rotterdam. This laboratory takes part in the standardization program of the WHO and is the Lipid reference laboratory for standardized cholesterol determinations in the Netherlands.

Data processing and statistical analyses

All analyses were performed using SAS-software, version 6.07. An adapted version of the 1993 computerized Dutch food composition table [9] was used to calculate energy and nutrient intakes. For those items in the food frequency questionnaire which were at a higher aggregation level than the foods in the food composition table, a weighted mean nutrient composition was calculated. The weights were derived from the database of the Dutch National Food Consumption Survey 1987/1988 [10]. Vitamin supplements used by 22% of the study population were not taken into account in calculating vitamin intakes. Total protein intake (g) was calculated from 24-h urinary nitrogen excretion (N_{exc} in g) by the formula $6.25*(N_{exc}/0.81)$ [11]. Urine sampling was considered incomplete if a subject reported that one or more of the collections was incomplete (n=28), or when less than four 24-h collections were made (n=6). BMR was estimated using standard formulas based on age, gender, height and weight [12]. One kg was subtracted from measured weight to correct for the weight of clothes.

Statistical analyses were performed on $\log_e$-transformed values, since nutrient intake and biomarker variables were generally skewed towards higher values. A weighted mean was calculated for the single 24-h recall nutrient intakes, with a weight of two for Saturdays and Sundays (which were recalled once by each person) and a weight of one for the other days (which were on average recalled twice by each person). Mean values for biomarker variables were calculated over the four repeats. Nutrient variables were adjusted for energy intake using the residual method [13].

Pearson correlation coefficients between the nutrient intakes assessed by means of the first and second and the first and third food frequency questionnaires were used to evaluate the 6 and 12-month reproducibility for ranking subjects according to nutrient intake. Pearson correlations between nutrient intakes calculated from the food frequency questionnaire and those assessed by reference methods were used to express the relative validity for the ranking of individuals according to nutrient intake. Since serum cholesterol is known to affect α-tocopherol levels [14], partial correlation coefficients adjusted for serum cholesterol were calculated. In addition, correlations between serum α-tocopherol and vitamin E intake were calculated for the subgroup of non-supplement users and between serum ß-carotene and ß-carotene intake for the subgroup of non-smokers [14]. Because 12 repeated 24-h dietary recalls and four repeated biomarker measurements may not be enough to account for all intraindividual variation, we deattenuated the correlation coefficients by multiplying them by the factor

$(1 + (\sigma^2_w/\sigma^2_b)/n)^{0.5}$, where n is the number of repeated measures, σ^2_w is the intraindividual variance, and σ^2_b is the interindividual variance [2]. The variance components were estimated by random effects models with the recall or biomarker variable as the dependent variable and subject number as the independent variable. In the models using 24-h recall variables a weight of two was again given to Saturdays and Sundays and a weight of one to the other days.

To assess the relative validity for the absolute intake on the group level, a linear measurement error model was considered; it is assumed that bias consists of a constant part which is the same for each individual and a proportional part, i.e. a bias which is correlated with the level of true intake [15]. Linear regression models were constructed, with the food frequency questionnaire nutrient intake as dependent variable and the reference method nutrient intake as independent variable. Two-sided t-tests were used to test whether regression coefficients differed from one (proportional bias), and to test whether the mean difference between intake assessed by the food frequency questionnaire and the reference method differed significantly from zero (constant bias).

Results

Sex-specific mean energy and nutrient intakes and their coefficients of variation (no transformation applied) as assessed by the 24-h recalls and the first food frequency questionnaire are shown in table 1. The observed mean intakes exhibit reasonable agreement with those reported by the Dutch National Food Consumption Survey 1992 [16].

In general, 6-month reproducibility for the ranking of subjects according to nutrient intake was better than 12-month reproducibility, and reproducibility for males was better than that for females (table 2). For men, all Pearson correlation coefficients between the nutrient intakes of the repeated questionnaires were 0.7 or higher, while for women the correlation coefficients for retinol, ß-carotene, and vitamin E were below 0.7. When non-drinkers were excluded, correlation coefficients for alcohol intake were 0.89 and 0.77 for men (n=54) and 0.87 and 0.90 for women (n=45). Adjustment of the nutrient intake for total energy intake resulted in lower 6-month reproducibility for men: the median correlation coefficient decreased from 0.83 to 0.74. Reproducibility for women and 12 month reproducibility for men were not appreciably altered.

Table 1. *Mean values and coefficients of variation for daily energy and nutrient intakes estimated by means of 12 24-h dietary recalls and the Dutch EPIC food frequency questionnaire (FFQ) in comparison with the Dutch National Food Consumption Survey 1992 (DNFCS)*[a]

Nutrient	Males (n=63)					Females (n=58)				
	recalls		FFQ		DNFCS	recalls		FFQ		DNFCS
	mean	cv%[b]	mean	cv%	mean	mean	cv%	mean	cv%	mean
Energy (Mj)	11.3	19.7	11.6	28.0	11.3	7.5	20.3	8.0	22.2	8.2
Energy (Kcal)	2701	19.7	2773	28.0	2709	1798	20.3	1894	22.2	1954
Protein (g)	97	18.5	99	25.5	98	70	19.2	72	21.1	76
Fat (g)	104	28.2	116	32.3	113	71	27.4	81	29.0	83
Carbohydrates (g)	295	25.4	298	34.0	290	198	23.1	202	27.5	211
Alcohol (g)	26	95.6	19	114.4	19	11	103.6	10	139.1	8
Dietary fibre (g)	18.8	26.7	20.3	32.2	18	14.2	25.2	15.5	27.0	15
Retinol (mg)	0.66	75.2	0.74	46.2	0.83[c]	0.45	64.3	0.55	66.6	0.68[c]
ß-carotene (mg)	1.62	43.9	1.88	48.0		1.22	48.7	1.51	40.3	
Vitamin C (mg)	92	38.4	113	45.9	69	81	42.8	103	38.0	78
Vitamin E (mg)	16	32.4	19	41.6	-	11	32.4	14	29.5	-
Energy %										
Protein	14.5	14.9	14.5	16.9	14.8	15.7	14.7	15.2	13.3	16.1
Fat	34.2	16.7	37.4	12.7	37.5	35.2	12.9	37.9	13.9	37.6
Carbohydrates	43.7	15.7	42.7	14.9	43.0	44.2	12.8	42.5	17.5	43.5
Alcohol	7.0	92.4	4.9	116.7	4.9	4.2	106.5	3.8	137.2	2.8

[a] Reference 16; For men compared with results for age-group 22-50 years, for women compared with a weighted average of results for age-groups 22-50 years (weight 0.6) and 50-65 years (weight 0.4).
[b] cv% = 100 * standard deviation / mean ; [c] Retinol equivalents: retinol + 1/6 ß-carotene

Expressed as Pearson correlation coefficients, the median validity relative to the 24-h recalls was 0.59 for men and 0.58 for women. The highest correlation coefficients were found for alcohol and the lowest for ß-carotene (table 3). Relative validity was also assessed for components of fat, protein, and carbohydrates, the results being similar to those for the main macronutrients (not shown). After excluding non-drinkers, the correlation for alcohol was 0.74 for males (n=54) and 0.87 for females (n=45). The median relative validity for energy-adjusted nutrients among men was 0.55, among women 0.59. For the second and the third food frequency questionnaires the median crude correlation coefficients were 0.63 and 0.58, respectively, for men, and 0.52 and 0.57, respectively, for women.

Table 2. *Reproducibility at 6 and 12 months for energy and nutrient intakes estimated by the Dutch EPIC food frequency questionnaire, expressed as Pearson correlation coefficients*[a]

Nutrient	Males (n=63)				Females (n=58)			
	crude		energy-adjusted[b]		crude		energy-adjusted	
	6 mth	12 mth	6 mth	12 mth	6 mth	12 mth	6 mth	12 mth
Energy	0.90	0.79	-	-	0.80	0.75	-	-
Protein	0.86	0.70	0.82	0.73	0.75	0.71	0.76	0.70
Fat	0.83	0.78	0.73	0.64	0.77	0.76	0.72	0.80
Carbohydrates	0.91	0.76	0.75	0.72	0.85	0.83	0.87	0.89
Alcohol	0.94	0.89	0.94	0.89	0.93	0.94	0.93	0.94
Dietary fibre	0.85	0.73	0.82	0.73	0.81	0.71	0.75	0.76
Retinol	0.81	0.74	0.74	0.68	0.66	0.66	0.63	0.60
ß-carotene	0.79	0.78	0.71	0.75	0.61	0.59	0.57	0.62
Vitamin C	0.78	0.74	0.72	0.75	0.76	0.70	0.74	0.71
Vitamin E	0.81	0.73	0.72	0.64	0.66	0.67	0.46	0.63

All variables were $\log_e$-transformed before analysis; [a] 95% confidence intervals for N=60 are -0.16-0.35 for r=0.1; -0.06-0.43 for r=0.2; 0.05-0.51 for r=0.3; 0.16-0.59 for r=0.4; 0.28-0.67 for r=0.5; 0.41-0.74 for r=0.6; 0.54-0.81 for r=0.7; 0.69-0.88 for r=0.8; 0.84-0.94 for r=0.9; [b] Energy-adjustment was performed according to the residual method [13]

Table 3. *Pearson correlation coefficients between daily intake of nutrients assessed by the Dutch EPIC food frequency questionnaire and by 12 24-h dietary recalls.*

Nutrient	Males (n=63)				Females (n=58)			
	non-adjusted		energy-adjusted[a]		non-adjusted		energy-adjusted	
	crude	deatt[b].	crude	deatt.	crude	deatt.	crude	deatt.
Energy	0.71	0.77	-	-	0.58	0.62	-	-
Protein	0.61	0.68	0.62	0.71	0.51	0.56	0.59	0.67
Fat	0.69	0.74	0.57	0.61	0.58	0.63	0.57	0.63
Carbohydrates	0.72	0.75	0.71	0.74	0.66	0.69	0.72	0.76
Alcohol	0.83	0.87	0.82	0.85	0.90	0.94	0.84	0.87
Dietary fibre	0.51	0.56	0.55	0.61	0.67	0.75	0.65	0.74
Retinol	0.57	0.61	0.26	0.29	0.57	0.62	0.54	0.62
ß-carotene	0.26	0.34	0.23	0.32	0.35	0.47	0.23	0.31
Vitamin C	0.39	0.45	0.37	0.43	0.58	0.69	0.61	0.71
Vitamin E	0.57	0.63	0.53	0.58	0.44	0.48	0.35	0.41

All variables were $\log_e$ transformed before analyses; [a] energy adjustment according to the residual method [13]; [b] corrected for intraindividual variation in 24-h dietary recalls [2]

Deattenuation of the correlation coefficients to correct for the lack of precision in individual mean recall values due to intraindividual variation increased the relative validity for ß-carotene by a factor 1.32 to 0.34 for men and by a factor 1.35 to 0.47 for women. Deattenuation factors for other nutrients were smaller. The median deattenuated crude correlation coefficients between nutrient intakes estimated by means of the first dietary questionnaire and by 24-h recalls became 0.66 for men and 0.63 for women.

Table 4. *Pearson correlation coefficients between intake of vitamin E, ß-carotene, and protein assessed by the Dutch EPIC food frequency questionnaire and by biomarkers*

	Males (n=63)		Females (n=58)	
Biomarker	crude	deattenuated[a]	crude	deattenuated
Serum α-tocopherol[b]	0.20	0.21	0.07	0.07
Serum α-tocopherol[b], no supplement use[c]	0.23	0.24	0.15	0.15
Serum ß-carotene	-0.16	-0.17	0.13	0.14
Serum ß-carotene, non-smokers[d]	-0.16	-0.17	0.11	0.12
Urinary nitrogen	0.37	0.43	0.45	0.50
Urinary nitrogen, complete[e]	0.47	0.56	0.53	0.58
Energy-adjusted[f] questionnaire nutrient intakes:				
Serum α-tocopherol[b]	0.29	0.30	0.14	0.14
Serum α-tocopherol[b], no supplement use[c]	0.32	0.33	0.13	0.13
Serum ß-carotene	-0.14	-0.15	0.17	0.18
Serum ß-carotene, non-smokers[d]	-0.08	-0.08	0.17	0.18
Urinary nitrogen	0.41	0.48	0.53	0.59
Urinary nitrogen, complete[e]	0.47	0.56	0.63	0.69

All variables were $\log_e$-transformed before analyses; [a] corrected for intraindividual variation in biomarker values [2]; [b] partial correlation coefficients corrected for serum total cholesterol level; [c] n=56 for men, and n=50 for women; [d] n=41 for men, and n=39 for women; [e] n=46 for men, and n=43 for women; [f] energy adjustment according to the residual method [13]

Pearson correlation coefficients between biomarker variables and nutrient intakes as assessed by the dietary questionnaire are given in table 4. With the exception of protein intake among women, these correlation coefficients were lower than those for the intake assessed by the questionnaire versus that assessed by 24-h recalls. For ß-carotene a negative correlation between the serum concentration and intake estimated by the food frequency questionnaire was found for men. Correction for intraindividual

variation in biomarker values had the greatest effect on the correlation coefficient for urinary nitrogen. Correlation coefficients between biomarker values and energy-adjusted nutrient estimates were generally higher than those for unadjusted nutrient estimates.

Compared to the 24-h recall data a constant positive bias of more than 10% was observed for the intake of fat, alcohol, dietary fibre, and (pro)vitamins as assessed by the dietary questionnaire (table 5). Although statistically significant ($p<0.05$), the constant positive bias in energy and protein intake among women was smaller. For all of these nutrients, except alcohol intake among women, a significant proportional bias was also present. The combination of a positive constant bias and regression coefficients below one means that overestimation decreases with increasing intake. The median regression coefficient for women was lower than that for men (0.54 versus 0.71).

Table 5. *Mean daily nutrient intakes based on 24-h recalls and the Dutch EPIC food frequency questionnaire (ffq), and regression coefficients (ß) from regressing intake assessed by food frequency questionnaire on intake assessed by 24-h recalls*

Nutrient	men (n=63)			women (n=58)		
	recall mean	ffq mean	ß	recall mean	ffq mean	ß
Energy (Mj)	10.7	11.2	1.00	7.1	7.8**	0.55***
Energy (Kcal)	2570	2667	1.00	1701	1856	0.55
Protein (g)	92	96	0.85	66	70*	0.49***
Fat (g)	93	110***	0.74*	63	77***	0.52***
Carbohydrates (g)	277	282	0.94	187	194	0.75*
Alcohol (g)	3	7***	0.75***	1	3***	0.92
Dietary fibre (g)	17	19**	0.66*	13	15***	0.66**
Retinol (mg)	0.43	0.66***	0.56***	0.32	0.47***	0.64**
ß-carotene (mg)	1.02	1.71***	0.27***	0.73	1.40***	0.33***
Vitamin C (mg)	65	102***	0.32***	56	96***	0.45***
Vitamin E (mg)	14	18***	0.67*	9	13***	0.38***

All variables were $\log_e$ transformed before analyses. Means were transformed back.
Hypothesis testing: mean difference recall versus ffq $\neq 0$ and ß$\neq 1$ by t-test.
Two-sided p-values: *=p<0.05, **=p<0.01, ***=p<0.001.

When protein intake based on nitrogen excretion was taken as a reference method, constant bias and proportional bias were observed for both men and women

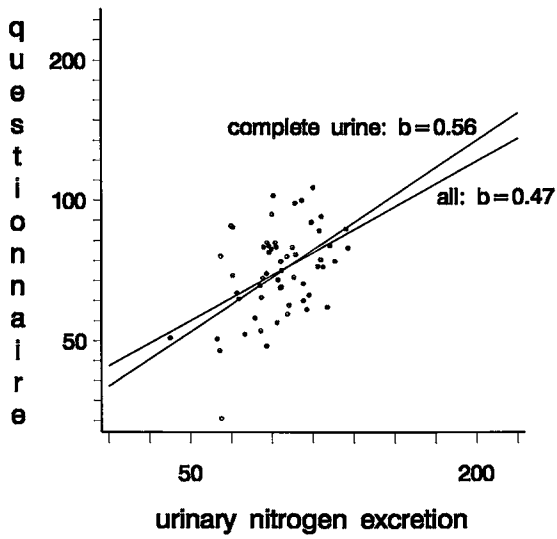
who turned in complete urine samples (figure 2). In contrast to the comparison with 24-h recalls the constant bias was negative, implying underestimation by the food frequency questionnaire. The mean ratio of the energy estimate from the questionnaire to BMR was 1.53 (sem 0.05) for men, and 1.39 (sem 0.04) for women. The ratio among women was significantly different ($p<0.05$) from the reference value of 1.55. This indication of underestimation of energy intake among women contradicts the result for energy in table 5 obtained with 24-h recalls as reference method.

Discussion

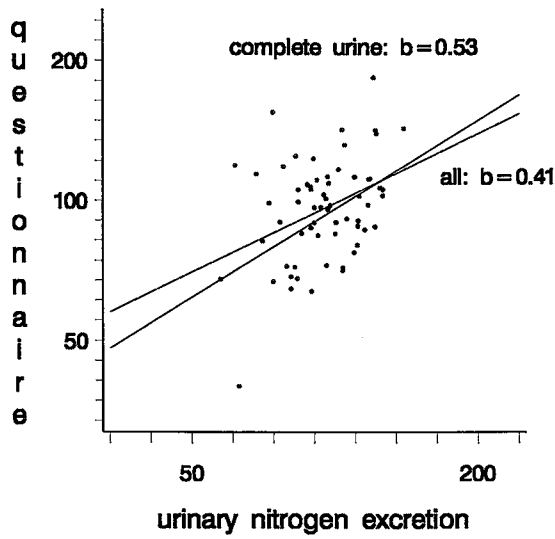
This study was conducted to evaluate the reproducibility and relative validity for a self-administered semi-quantitative food frequency questionnaire. The observed median crude correlation coefficient of 0.58-0.59 between the food frequency questionnaire and the main reference method compares favorably with those reported for several US questionnaires [17-19], and one Danish one [20]. The results of a Finnish questionnaire were very close to ours [21], while for a Dutch [22] and a Norwegian questionnaire [23] somewhat higher median correlation coefficients were obtained when the same nutrients were compared. For the present questionnaire as well as other questionnaires [17-23] the range of correlation coefficients with the reference method was wide and it seems virtually impossible to make a food frequency questionnaire that performs well for a long list of nutrients and food groups. Median reproducibility of the present questionnaire was better than that of similar food frequency questionnaires [17,19,21-23]. It should however be kept in mind that these comparisons are crude, due to differences in study populations, reference methods, and methodology. For instance in the studies mentioned above, dietary records were used as reference methods.

The response for this validation study was 25% and a selection of the participants towards more health conscious people is likely. Therefore, the study population in this validation study is probably better able to describe their food habits than the average member of the Dutch EPIC cohort.

The 24-h dietary recall method is conceptually different from the food frequency method, since it is open-ended, depends on short-term memory, and does not require the subjects to average out consumption frequencies and portion sizes over time. However, both methods have some sources of error in common, e.g. use of the same food composition table and subject-specific underreporting or overreporting. This will lead to artificially high correlation coefficients between the dietary recalls and the



females, n=43 for women with complete urine samples
mean protein estimate: questionnaire: 71 g, urinary N: 81 g



males, n=46 for men with complete urine samples
mean protein estimate: questionnaire: 92 g, urinary N: 99 g

Figure 2. *Protein intake (g) based on the food frequency questionnaire versus that calculated from 24-h urinary nitrogen excretion*

food frequency questionnaire. On the other hand, random errors in the 24-h dietary recalls which are not correlated with those in the questionnaire would tend to falsely lower correlations. We adjusted the correlation coefficients for part of this error due to day-to-day variation in intake. This increased the median correlation coefficients moderately from 0.59 to 0.66 for men and from 0.58 to 0.63 for women. Whether the correlation coefficients obtained after this correction are underestimates or overestimates of the true correlations remains unknown.

Biomarkers satisfy the criterion of errors which are independent of those of the questionnaire [24]. Urinary nitrogen has been proven to reflect protein intake accurately, since a correlation of about one was found in a metabolic ward situation. The only problems in less controlled situations are incompleteness of the urine samples and day-to-day variation in excretion [11]. For females, the difference between the relative validity for protein intake compared to the 24-h recalls (r=0.56) and that compared to nitrogen excretion (r=0.58) was very small after adjustment for day-to-day variation and exclusion of incomplete urine collections. For males, the difference was larger; we suspect that the correlation coefficient with the 24-h recalls (0.68) is an overestimation of true validity, while that with nitrogen excretion (0.56) is a closer approximation of the truth.

Relationships between serum concentrations of α-tocopherol and ß-carotene and their intake are inevitably confounded and attenuated by individual variations in absorption, availability and metabolism [25]. Further attenuation occurs because food composition data are of limited quality for vitamin E and ß-carotene [26], and because vitamin E intake was compared with the blood concentration of α-tocopherol alone. Measurement errors in the laboratory also contribute to artificially low correlation coefficients (CV ß-carotene 12.5%). Our poor results for serum α-tocopherol and ß-carotene can thus be interpreted as the lower limit of the true validity. This is supported by the observation that correlations between the serum antioxidants and 24-h recall estimates were also low (vitamin E: 0.38 for men and 0.22 for women; ß-carotene: 0.14 for men and 0.15 for women).

The correlation coefficient between the serum α-tocopherol level and vitamin E as estimated by means of the dietary questionnaire was 0.23 for male and 0.15 for female non-supplement users (adjusted for the serum cholesterol level). These correlation coefficients are low, but they fit well within the range of correlation coefficients for non-supplement users found for similar questionnaires [14,24,25,27-29].

Since insufficient information was available on vitamin content of supplements, we were not able to investigate relative validity for vitamin E intake from diet and supplements together, which tends to improve the correlation with reference methods [14]. The discrepancy of about 0.3 between the correlation coefficients for serum levels and those for 24-h recalls is large, which illustrates our inability to assess true validity for vitamin E intake assessed by means of the questionnaire.

The correlation coefficients between the serum ß-carotene level and ß-carotene intake assessed by means of the questionnaire were very low, i.e. negative for men and 0.13 for women. Even among non-smokers the correlations were not higher. This is contrary to expectations, since reports in literature describe correlation coefficients between 0.25 and 0.45 for non-smokers and correlations between 0 and 0.20 for smokers [14,27,28]. The correlation coefficients between ß-carotene intake assessed by the food frequency questionnaire and that assessed by 24-h recalls were higher but still poor, i.e. 0.34 for men and 0.47 for women after deattenuation for day-to-day variation. The relative validity of the food frequency questionnaire with respect to ß-carotene seems therefore limited. This is in accordance with our observation in the companion paper that the relative validity of vegetable intake, the main contributor to ß-carotene intake, is poor [4]. A second plausible explanation is the fact that some food items in the food frequency questionnaire, such as 'soup' and 'cabbage & kale', are heterogeneous with respect to ß-carotene composition. Furthermore, Dutch regulations do not allow ß-carotene fortification of foods, which means that many different foods each contribute a little to the total ß-carotene intake. This makes it difficult to estimate ß-carotene intake correctly by means of a questionnaire.

If in the Dutch EPIC cohort, measurement error for nutrient intake assessed by means of the food frequency questionnaire is random with respect to cancer outcome, then this will tend to bias measures of association towards the null value in most situations. Crude estimates of log relative risks for one unit difference in intake will be biased by a factor which is equal to the inverse of the regression coefficient representing proportional bias, multiplied by the square of the correlation coefficient between measured and true intake [15]. In this validation study we tried to gain an impression of the bias factor by estimating the proportional scaling and correlation coefficients for the nutrients of interest. A median bias factor of 0.5-0.6 as was found in this study would result in a true relative risk of 2 to be observed as between 1.4 and 1.5.

In practice, measurement error is often associated with other factors, such as body mass index (BMI), which also can be related to disease outcome. In this case the influence of measurement error on the relative risk can be in either direction. Indications for more frequent underreporting of energy intake by subjects with a higher BMI were also present in this dataset. Average BMI equalled 26.0 kg/m^2 for males (n=17) and 25.5 kg/m^2 for females (n=19) among underreporters, which were defined as persons with an energy intake based on the food frequency questionnaire below 1.2 times their predicted BMR [30]; for the other subjects mean BMI was lower, i.e. 24.9 and 24.0 kg/m^2, respectively.

Systematic bias is of importance for the intercohort analyses of the EPIC study. A systematic bias of nutrient intake as estimated by the food frequency questionnaire can only be evaluated by comparing that intake with an intake assessed by means of a 'gold' standard which is not subject to this type of bias. For the estimation of bias in protein intake assessed by the food frequency questionnaire we observed a discrepancy between results based on 24-h recalls (positive bias) and results based on 24-h urinary nitrogen excretion (negative bias), implying that the 24-h recalls underestimate protein intake more than the questionnaire. Among women, comparison of the energy intake based on the questionnaire with BMR and with 24-h recalls also led to contradictory conclusions. Clearly the 24-h recall method is not a gold standard for determining absolute intake at the group level and results should therefore be interpreted with caution.

Acknowledgements

The authors thank all subjects in the study for their time and effort. We are grateful to Karin Blank, Anneke Blokstra, Sigrid Braak, Anja Muusers and Lucie Viet for the dietary interviewing, Birgit Koopman-de Zeeuw for administrative support, and the laboratory technicians. We acknowledge Ester Goddijn, Annemarie Jansen, Gatske Obermann-de Boer, and Petra Peeters for their participation and advice during the planning and data collection phase.

This study was part of the EPIC project, and was supported financially by the Europe Against Cancer Program of the European Union and by the Ministry of Welfare, Health and Culture in the Netherlands.

References

1. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol 1992; 3 783-91
2. Willett W. Nutritional Epidemiology, Monographs in Epidemiology and Biostatistics. Vol. 15. New York, Oxford: Oxford University Press, 1990
3. Nelson M. Assessment of food consumption and nutrient intake. Past intake. In: Margetts B M, Nelson M (eds). Design concepts in nutritional epidemiology. Oxford, New York, Tokyo: Oxford University Press, 1991, pp 167-91
4. Ocké MC, Bueno de Mesquita HB, Goddijn HE, et al. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire and relative validity and reproducibility for food groups, in press
5. FAO/WHO/UNU. Report of a joint expert consultation. Energy and Protein requirements. WHO Technical Report Series No. 724. WHO: Geneva, 1985
6. Vatassery GT, Mortenson GA. Manual spectrophotometric and fluorometric determination of tocopherol in cerebrospinal fluid. Clin Chem 1972; 18: 1475-7
7. Craft NE, Brown ED, Smith JC. Effects of storage and handling conditions on concentrations of individual carotenoids, retinol, and tocopherol in plasma. Clin Chem 1988; 34: 44-8
8. Katterman R, Jaworek D, Möller G. Multicentre study of a new enzymatic method of cholesterol determination. J Clin Chem Clin Biochem 1984; 22: 245-51
9. Stichting NEVO. Nederlands voedingsstoffenbestand 1993. The Hague: Voorlichtingsbureau voor de Voeding, brochure 202, 1993
10. Hulshof KFAM, Van Staveren WA. The Dutch National Food Consumption Survey: design, methods and first results. Food Policy 1991; 16: 257-60
11. Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. Am J Clin Nutr 1985; 42: 1276-89
12. Schofield WN, Schofield C, James WPT. Basal metabolic rate _ review and predictions, together with an annotated bibliography of source material. Hum Nutr Clin Nutr 1985; 39C: (suppl 1), 4-96
13. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986; 124: 17-27
14. Stryker WS, Kaplan LA, Stein EA, Stampfer MJ, Sober A, Willett WC. The relation of diet, cigarette smoking, and alcohol consumption to plasma beta-carotene and alpha-tocopherol levels. Am J Epidemiol 1988; 127: 283-96
15. Kaaks R, Plummer M, Riboli E, Estève J, Van Staveren W. Adjustment for bias due to errors in exposure assessments in multicenter cohort studies on diet and cancer: a calibration approach. Am J Clin Nutr 1994; 59: (suppl) 245S-50S
16. Voorlichtingsbureau voor de Voeding. Zo eet Nederland, 1992. Resultaten van de Voedselconsumptiepeiling 1992. The Hague: Voorlichtingsbureau voor de Voeding, 1993
17. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985; 122: 51-65
18. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. Epidemiology 1990; 1: 58-64

19. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992; 135: 1114-26
20. Tjønneland A, Overvad K, Haraldsdóttir J, Bang S, Ewertz M, Jensen OM. Validation of a semiquantitative food frequency questionnaire developed in Denmark. Int J Epidemiol 1991; 20: 906-12
21. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. Am J Epidemiol 1988; 128: 655-66
22. Goldbohm RA, Van den Brandt PA, Brants HAM, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr 1994; 48: 253-65
23. Nes M, Frost Andersen L, Solvoll K, et al. Accuracy of a quantitative food frequency questionnaire applied in elderly Norwegian women. Eur J Clin Nutr 1992; 46: 809-21
24. Willett WC, Stampfer MJ, Underwood BA, Speizer FE, Rosner B, Hennekens CH. Validation of a dietary questionnaire with plasma carotenoid and α-tocopherol levels. Am J Clin Nutr 1983; 38: 631-9
25. Ascherio A, Stampfer MJ, Colditz GA, Rimm EB, Litin L, Willett WC. Correlations of vitamin A and E intakes with the plasma concentrations of carotenoids and tocopherols among american men and women. J Nutr 1992; 122: 1792-801
26. Romieu I, Stampfer MJ, Stryker WS, et al. Food predictors of plasma beta-carotene and alpha-tocopherol: validation of a food frequency questionnaire. Am J Epidemiol 1990; 131: 864-76
27. Bolton-Smith C, Casey CE, Gey KF, Smith WCS, Tunstall-Pedoe H. Antioxidant vitamin intakes assessed using a food-frequency questionnaire: correlation with biochemical status in smokers and non-smokers. Br J Nutr 1991; 65: 337-46
28. Coates RJ, Eley JW, Block G, et al. An evaluation of a food frequency questionnaire for assessing dietary intake of specific carotenoids and vitamin E among low-income black women. Am J Epidemiol 1991; 134: 658-71
29. Jacques PF, Sulsky SI, Sadowski JA, Phillips JC, Rush D, Willett WC. Comparison of micronutrient intake measured by a dietary questionnaire and biochemical indicators of micronutrient status. Am J Clin Nutr 1993; 57: 182-9
30. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. derivation of cut-off limits to identify under-recording. Eur J Clin Nutr 1991; 45: 569-81

Chapter 7

Biochemical markers as an additional measurement in dietary validity studies:

Application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition

Marga Ocké, Rudolf Kaaks

Am J Clin Nutr, in press

Abstract

The validity coefficient of dietary questionnaire measurements can be estimated from a triangular comparison between questionnaire, reference and biomarker measurements in a validity study using the method of triads. The method assumes that the measurements are linearly related to true intake and have independent random errors. We applied the method of triads to examples from the European Prospective Investigation into Cancer and Nutrition. In some examples 'Heywood cases' occurred, i.e., the estimated validity coefficients were >1, or the validity coefficients were not estimable. Such results are caused by random sampling fluctuations or violation of the model assumptions. One possible violation is a positive correlation between the random errors of questionnaire and reference measurements. We also demonstrated the use of a bootstrap method to estimate confidence intervals for the validity coefficients. Validity studies with several hundred subjects, more accurate biochemical indicators of dietary intake, or both are needed to estimate validity coefficients precisely and avoid complications occurring with the bootstrap method.

Introduction

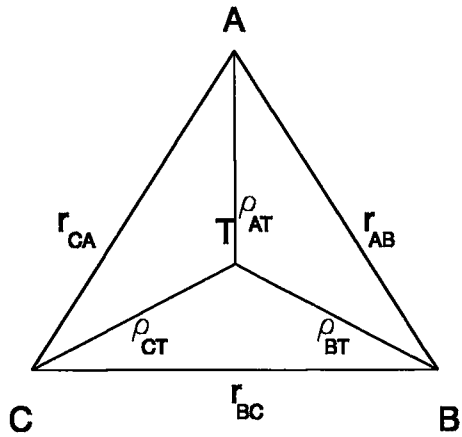
The aim of most dietary validity studies is to estimate the correlation between dietary questionnaire data and the subjects' true habitual intake levels, the validity coefficient [1]. This coefficient is usually estimated from the correlation with the mean values obtained from multiple food records or 24-hour recalls, with correction for the attenuating effects due to random variations in these reference measurements. If the random errors of questionnaire and reference measurements are positively correlated, the validity coefficient is overestimated. If there is a positive correlation between the random errors of multiple reference measurements, the coefficient is underestimated [2]. Without non-questionnaire information it is impossible to predict which of the two possible biases - overestimation or underestimation - will predominate.

It is for this reason that the use of biochemical markers is increasing in dietary validity studies. The advantage of marker assessments is that the random errors occurring with their is are likely to be truly independent of those in both questionnaire measurements and reference measurements such as food records or 24-h recalls. In most published dietary validity studies, the additional information obtained from the comparison with biochemical marker results has been reported as a separate, additional correlation coefficient between the questionnaire and marker measurements. This correlation, even if sometimes rather low, was considered as evidence that the questionnaire measurements must have at least some level of validity (see reviews by van 't Veer *et al.* [3] and Willett [4]).

Kaaks [5] has described a triangular comparison between questionnaire, reference and biochemical marker measurements that can be used to obtain a quantitative estimate of the questionnaires' validity coefficient. This approach, called the method of triads [6], assumes that correlations between the three measurements are explained entirely by the fact that they all are linearly related to the true intake levels and that their random measurement errors are mutually independent. The method of triads, a basic estimating technique in factor and path analysis, is based on fitting a theoretical to an observed correlation matrix.

In this paper we illustrate application of the method of triads by using selected examples from validity studies conducted during the pilot phase of the European Prospective Investigation into Cancer and Nutrition (EPIC) [7]. We also demonstrate the use of a bootstrap method to obtain confidence intervals for the estimated validity

coefficients and discuss the strengths and limitations of these methods in practical situations.



$$r_{AB} = \rho_{AT} * \rho_{BT}$$
$$r_{BC} = \rho_{BT} * \rho_{CT}$$
$$r_{CA} = \rho_{AT} * \rho_{CT}$$

$$\rho_{AT} = \sqrt{r_{AB} * r_{CA} / r_{BC}}$$
$$\rho_{BT} = \sqrt{r_{AB} * r_{BC} / r_{CA}}$$
$$\rho_{CT} = \sqrt{r_{CA} * r_{BC} / r_{AB}}$$

Figure 1. *Graphical representation of the method of triads. T=true intake variable; A,B,C=measurements of type A, B, or C; ρ=validity coefficient; r=sample correlation*

The method of triads: theoretical examples

Before examining the selected examples from the EPIC-study, we will briefly describe the quantitative relations between the measurements' validity coefficients (ρ) and the correlations between the three types of measurements expected to be observed in a validity study (=sample correlations, r). We will refer to the three hypothetical types of measurements as 'A', 'B', and 'C'. As the equations in figure 1 indicate, when all three measurement types have a high validity coefficient, (e.g., $\rho_{AT} = \rho_{BT} = \rho_{CT} = 0.80$), the sample correlations between the measurements are also expected to be relatively high ($r_{AB} = r_{BC} = r_{CA} = 0.64$). If one of the three types of measurements is relatively inaccurate (e.g., $\rho_{AT} = 0.30$, $\rho_{BT} = \rho_{CT} = 0.80$), however, two of the three sample correlations between measurements are expected to be low ($r_{AB} = r_{CA} = 0.24$, $r_{BC} = 0.64$). Inaccuracy of two of the three measurement types (e.g., $\rho_{AT} = 0.80$, $\rho_{BT} = \rho_{CT} = 0.30$) will cause all three expected sample correlations to be weak ($r_{AB} = r_{CA} = 0.24$, $r_{BC} = 0.09$). When all three types of measurements are inaccurate (e.g. $\rho_{AT} = \rho_{BT} = \rho_{CT} = 0.30$), all three sample correlations are expected to be weak ($r_{AB} = r_{BC} = r_{CA} = 0.09$).

Reasoning from the opposite approach, it is clear from the equations in figure 1 that the estimated validity coefficient for a type of measurements is always equal to or greater than the sample correlations between that type of measurement and the other two. Thus, if all three sample correlations are relatively high (eg,> 0.70), it can be concluded even without further calculations, that all three measurement types are expected to have a validity coefficient rather close to 1. If all three sample correlations are low (eg <0.3), it does not necessarily mean that all three types of measurements are inaccurate. One of the validity coefficients may still be high (as shown above). However, small differences in sample correlations that are low may result in rather large differences in the estimated validity coefficients of the three types of measurements.

The method of triads: selected examples with data from the EPIC study

In our examples from the EPIC validity studies (summarized in figure 2), the measurements to be evaluated were obtained by means of a semiquantitative food frequency questionnaire (Q). Reference measurements were based on the mean values from 12 24-h recalls (R) obtained at monthly intervals after (a first) administration of the questionnaire. Depending on the specific example, the marker value (M) was the mean of four quarterly measurements of 24-h urinary nitrogen excretion (a marker of protein intake) or serum concentrations of ß-carotene, vitamin C or total cholesterol.

In example 1 (measurements of protein intake), the sample correlations were moderately high, ranging from 0.45 to 0.59. The estimated validity coefficients were therefore also all relatively elevated, ranging from 0.63 for the questionnaire to 0.82 for the reference measurements. In example 2 (measurements of ß-carotene intake), the observed correlations between the methods were considerably lower than in example 1, and there was a corresponding decrease in two of the three estimated validity coefficients. However, the estimated validity coefficient of one of the three measurements was still relatively high (ρ_{RT} = 0.58).

In neither of the first two examples were there any special observations or complications in obtaining point estimates of the three validity coefficients with use of the method of triads. In example 3 (measurements of vitamin C intake), however, one of the estimated validity coefficients was slightly higher than 1 (ρ_{RT} = 1.01).

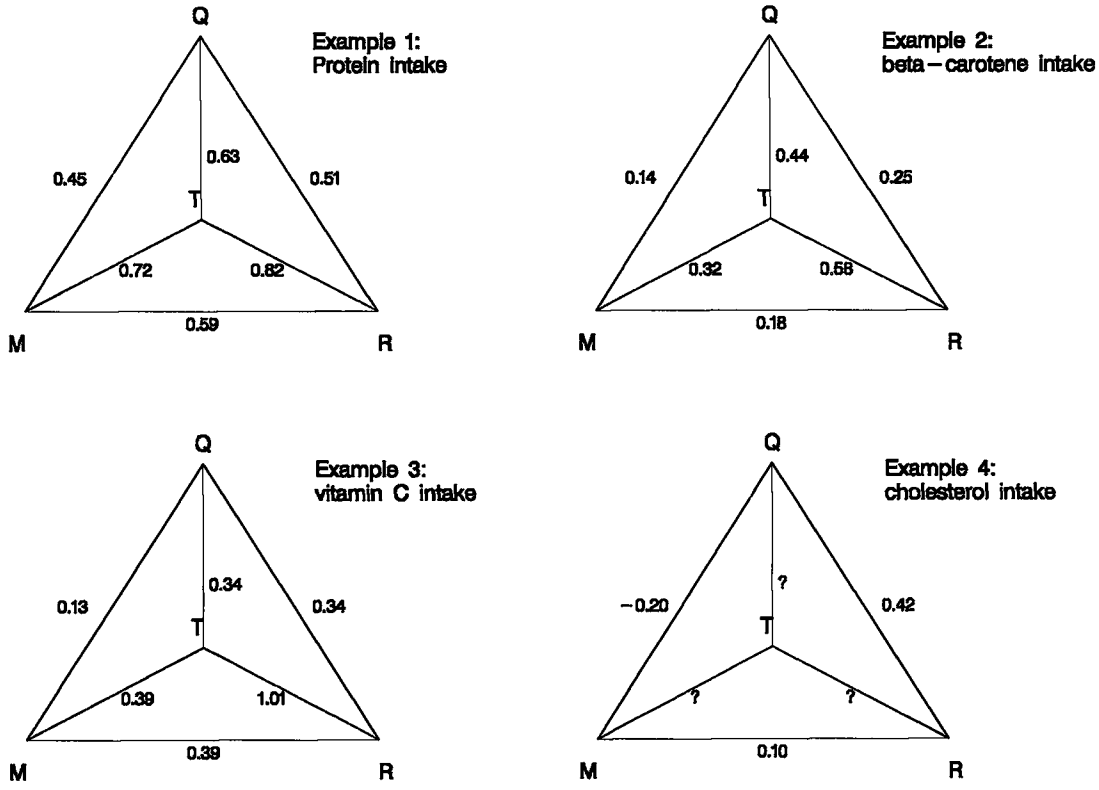


Figure 2. *Sample correlations and validity coefficients for triangular comparisons between questionnaire (Q), 24-h recall (R), and biochemical marker (M) measurements in examples from the European Prospective Investigation into Cancer and Nutrition*

An estimated validity coefficient >1 seems an anomalous finding since, by definition, correlation coefficients must be between -1 and 1. In the context of factor analysis, however, this kind of outcome is quite common, and is known as a Heywood case [8].

When the method of triads is employed, Heywood cases arise if the product of two of the three sample correlations is larger than the third. There are two different explanations for the occurrence of Heywood cases. The first is that random sampling fluctuations are present in the observed correlations between measurements. If, for instance, the validity coefficient of the reference measurements in example 3 was in reality just below 1, relatively small variations in the sample correlations may have led to the estimated coefficient of 1.01. For example, had the observed correlation r_{MQ} been equal to 0.14 rather than 0.13, the validity coefficient of the reference measurements would have been estimated as 0.97. To the extent that random fluctuations are the true explanation for the occurrence of the Heywood case, estimated validity coefficients with values >1 may be perfectly acceptable.

The second explanation for the occurrence of Heywood cases is that one or more of the underlying model assumptions (linear relations with truth and independence of random errors between the measurements) is violated. In such situations, the estimated validity coefficients are biased. For example, a positive correlation between the random errors of questionnaire and reference measurements would produce validity coefficients that are overestimated for the questionnaire and reference measurements and underestimated for the biochemical marker measurements.

In example 4 (measurements of cholesterol intake), a more serious complication occurred, since one of the three sample correlations (r_{MQ}) is negative. In this situation the method of triads cannot provide estimates of the validity coefficients because it would require taking the square root of a negative value. Random sampling fluctuations seem to be the most likely explanation for such a situation, which is also a Heywood case. Thus, if the true validity coefficient of the marker measurements is close to 0, there is a high probability that a negative sample correlation with one of the other two measurements will be observed in the validity study. This may have happened in the example with total serum cholesterol, which is a notoriously poor indicator of cholesterol intake [4]. Increasing the sample size of the validity study decreases the amplitude of sampling fluctuations. Theoretically, a negative correlation between, the random errors of questionnaire and biochemical marker measurements may also explain such a Heywood case.

Confidence intervals for the estimated validity coefficients: the bootstrap method

In studies on the accuracy of dietary questionnaire measurements, not only should the validity coefficient be estimated without bias, but the level of precision with which these estimates are obtained should be evaluated. A nonparametric approach that does not presume any knowledge about the theoretical probability distribution of the estimated validity coefficient is the bootstrap method [8,9]. This method involves the repeated drawing of samples from the group of individuals in whom measurements have been observed. The sampling is done with replacement, which allows each case to be drawn once, more than once, or not at all in each of the samples. The bootstrap samples are usually chosen to be of the same size as the number of individuals in the data set.

By applying the method of triads to each bootstrap sample, researchers can obtain empirical distributions of the three estimated validity coefficients. Efron and Gong [9] have shown that, in general, these empirical distributions will approximate the true theoretical probability distributions of the estimated validity coefficients. The empirical distributions can therefore be used to determine confidence intervals of the estimated variables. A bootstrap routine that uses the method of triads for estimation of the validity coefficients in each bootstrap sample can be programmed quite easily using any basic programming language. The bootstrap program used for the example in this paper was written in GLIM [10] (Appendix A). For our examples with data from the EPIC study, 200 bootstrap samples were drawn; this number is sufficient for most practical purposes [8].

Table 1 shows approximate probability limits obtained with the bootstrap method for the examples 1, 2 and 3. There are about 60 subjects in these examples. No estimates of probability limits are given for example 4, since the validity coefficients itself could not be estimated by the method of triads. The empirical cumulative distribution of the estimated validity coefficient of the questionnaire measurements in example 1 is shown in figure 3. The 95% probability limits for the estimated validity coefficient can be read from this empirical distribution and is 0.33-0.82. The probability interval for the validity coefficient of the multiple 24-hour recalls (0.63-1.07) includes 1, indicating the presence of Heywood cases in some of the bootstrap samples. This observation illustrates the effects of random sampling fluctuations.

Table 1. *Sample correlations and estimated validity coefficients with approximate probability limits of questionnaire (Q), 24-h recall (R) and biochemical marker (M) measurements in examples from the European Prospective Investigation into Cancer and Nutrition*[a].

	Example 1 protein intake (n=61)	Example 2 ß-carotene intake (n=61)	Example 3 vitamin C intake (n=56)
Sample correlations			
r_{QR} (95% ci)	0.51 (0.30-0.68)	0.25 (-0.00-0.47)	0.34 (0.08-0.55)
r_{RM} (95% ci)	0.59 (0.40-0.73)	0.18 (-0.08-0.41)	0.39 (0.14-0.49)
r_{MQ} (95% ci)	0.45 (0.22-0.63)	0.14 (-0.12-0.38)	0.13 (-0.14-0.38)
Validity coefficients			
% excluded[b]	0	20	17
ρ_{QT} (95% ci)	0.63 (0.33-0.82)	0.44 (0.09-1.25)	0.34 (0.07-0.94)
ρ_{RT} (95% ci)	0.82 (0.63-1.07)	0.58 (0.17-1.86)	1.01 (0.38-2.54)
ρ_{MT} (95% ci)	0.72 (0.50-0.87)	0.32 (0.07-0.83)	0.39 (0.08-0.79)

[a] 95% CIs in parentheses; [b] Percentage of bootstrap samples that were excluded, because one or three of the sample correlations were negative.

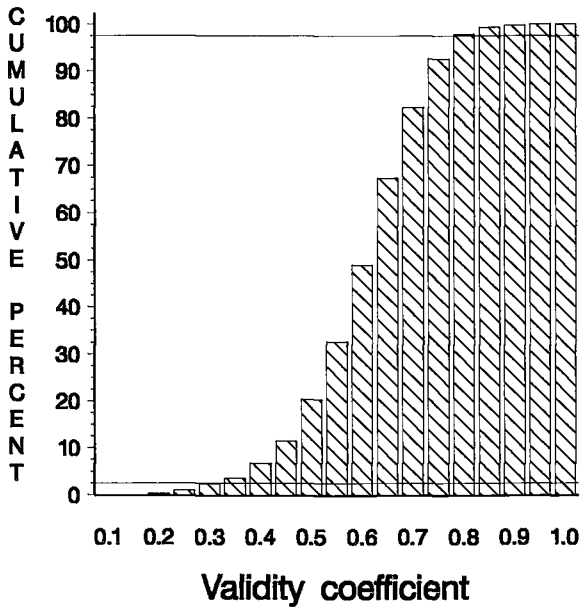


Figure 3. *Empirical cumulative distribution of the validity coefficient of questionnaire measurements obtained by the bootstrap method. Data from example 1 (protein intake)*

The approximate probability limits for example 2 are considerably wider than those in example 1. Since both examples are based on the same number of subjects, it can therefore be concluded that the probability intervals of the validity coefficients become wider when the sample correlations are poorer.

There were no insurmountable problems in obtaining point estimates of the validity coefficients in examples 1, 2 and 3. However, in examples 2 and 3 estimation of the validity coefficients was impossible in 20% and 17%, respectively, of the bootstrap samples because negative sample correlations between measurements were obtained. As a result, the approximate probability intervals for examples 2 and 3 in table 1 are not really 95% confidence intervals, since they are based on <85% of the bootstrap samples. Clearly, such a situation is probable if one or more of the three sample correlations in the validity study has a 95% confidence interval that includes 0. This was the case in example 2 for all three sample correlations and in example 3 for the correlation between the food frequency questionnaire and the biochemical marker. Increasing the sample size of validity studies and using more accurate reference and marker measurements, will decrease the likelihood of negative sample correlations occurring when the bootstrap method is used.

Correlated errors of questionnaire and reference measurements

The method of triads assumes independence of random errors between the three types of dietary measurements. In practice, this model requirement should determine the choice of the types of measurements with which questionnaire measurements are compared in a validity study. Unfortunately, it may be difficult to find three types of dietary intake measurements that are different enough to assume *a priori* that their random errors are mutually independent. For the examples presented in this paper, we feel confident that the random measurement errors of the biochemical marker data were truly independent of those of questionnaire measurements and 24-hour recalls. However, because questionnaire and 24-hour recall measurements may have some sources of error in common [3,4], we cannot rule out the presence of a real positive correlation between their errors. If such a correlation is the only violation of the model assumptions, the method of triads will overestimate the validity coefficients of questionnaire and reference measurements. These estimates may therefore be most prudently interpreted as upper limits for the true validity coefficients [5].

On the other hand, the observed correlations r_{MQ} and r_{RM} can be considered as estimated lower limits of the validity coefficients. If replicate marker measurements are available, the observed correlations r_{MQ} and r_{RM} should preferably be corrected for the attenuating effects caused by within-subject random errors in the marker measurements. The estimated lower and upper limits of the validity coefficients would thus be given as:

$r_{MQ} < \rho_{QT} < \rho_{QT(triad)}$

$r_{RM} < \rho_{RT} < \rho_{RT(triad)}$

Maximum of r_{MQ} and $r_{RM} < \rho_{MT} < 1.0$

Interpreted this way, the questionnaire measurements in example 1 of the EPIC study data (protein intake) appear reasonably accurate, with a lower limit for the validity coefficient of 0.45, and an upper limit of 0.63. In examples 2 and 3 (beta-carotene and vitamin C), the questionnaire measurements appear to be less accurate, with validity coefficients in the estimated ranges of 0.14-0.44, and 0.13-0.34, respectively. It should be noted, however, that the lower limits for examples 2 and 3 are underestimates because of the probably low validity coefficients for the biochemical markers.

Discussion

We have illustrated the practical application of the method of triads, an elementary factor analysis approach, in dietary validity studies based on the comparison between questionnaire measurements, multiple 24-h recalls, and a biochemical indicator of diet. The same estimation of the validity coefficients through triangular comparison is also possible with use of a structural equations model approach [2]. An advantage of the method of triads is that it requires no special software for latent variable analysis. It can be applied with even a simple pocket calculator, starting from sample correlations between the three different types of measurements.

As we have shown, the interpretation of results obtained by the method of triads can be complicated by the occurrence of Heywood cases. These cases correspond to two slightly different types of situations, ie, those in which one of the validity coefficients is estimated to have a value > 1; and those in which no estimation is possible by means of the method of triads because one of the sample correlations is negative. For both types of situations, the factor analytical methods discussed here can be extended to include additional constraints for variable estimates, so that none of the

estimated validity coefficients will exceed a limiting range of theoretically acceptable values, e.g., 0-1.

In analyses in which variable constraints are introduced, there will not always be a perfect fit of the observed to the theoretical correlation matrix. However, a goodness-of-fit statistic may (in theory) be computed from the difference between theoretical and observed correlation matrices and used to evaluate whether there may be a gross violation of model assumptions. The parameter constraints are incorporated easily with use of special computer programs for latent variable analysis, such as *LISREL* (SPSS, Chicago)[11], *SAS-CALIS* (SAS Institute Inc, Cary, NC)[12], or *EQS* (BMDP statistical Software Inc, Los Angeles, CA)[13]. The advantage of these programs is that they can also be used to analyze more complex latent variable models. For example, information on the completeness of the 24-h urine collections can be included, as was done by Plummer and Clayton [14]. In addition, models with data for males and females together stratified according to sex, or models stratified according to smoking status, can be analyzed. Because our aim was to present a method that can be applied without sophisticated software, we decided to use unrestricted analysis and to identify extreme Heywood cases by simple evaluation of the estimated validity coefficients.

For estimating confidence intervals, we followed the recommendations of Dunn [8], who proposed that bootstrap methods be used as the general approach for evaluating the precision of reliability estimates. Because the bootstrap method is basically nonparametric, it requires no specific knowledge about the theoretical probability distributions of the estimated validity coefficients and the population distributions of the measurements (such as normality). In the structural equations model approach, confidence intervals of the estimated coefficients can also be obtained with use of parametric formulas. The confidence intervals in our examples, with sample size of approximately 60, were relatively wide. In view of the often rather low sample correlations between measurements, particularly for some biochemical markers, the commonly used sample size of dietary validity studies - generally not larger than 100 to 200 individuals - are in many situations insufficient to estimate the validity coefficients with reasonable precision.

Two of our examples had relatively high proportions of bootstrap samples with negative sample correlations. A large number of Heywood cases arising during bootstrap sampling can be assumed to indicate either a lack of precision with which

estimates were obtained or a serious violation of model assumptions. Larger sample sizes are needed to reduce the probability that sampling fluctuations will lead to Heywood cases in either the original data set or in a high proportion of bootstrap samples. Nonetheless, it should be kept in mind, however, that even in large validity studies, Heywood cases can occur as a result of relatively small sampling fluctuations if the validity coefficient of one of the measurements is very close to either 1 or 0.

Conclusion

Validity studies with biochemical markers as additional reference assessments allow estimation of the validity coefficient of the questionnaire measurements by means of the method of triads. However, because the questionnaire and the main reference measurements (often based on food records or 24-h recalls) may have positively correlated random errors, this coefficient may be overestimated. In such situations, the marker measurements allow expression of a range for the questionnaires' validity coefficient, with the (deattenuated) sample correlation between the questionnaire measurements and the marker measurements as the lower limit and the estimate obtained by the method of triads as the upper limit. Confidence intervals for the validity coefficient can be easily obtained with use of the bootstrap method. This method performs well only when the sample size of the validity study consists of several hundreds of subjects or the correlations between measurements are high. These conditions must also be met for the range of validity coefficients to be estimated with reasonable precision.

Acknowledgements

We thank Elio Riboli, Francoise Clavel, Heiner Boeing, Bertine Collette and Daan Kromhout for allowing us to use the data of the France, German, and Dutch EPIC-validity studies as examples. Corinne Casagrande for help with programming the bootstrap method in *GLIM*; and Bas Bueno de Mesquita, Daan Kromhout and Wija van Staveren are thanked for helpful comments and suggestions.

References

1. Armstrong BK, White E, Saracci R, Principles of exposure measurement in epidemiology. Oxford: Oxford University Press, 1992
2. Kaaks R, Riboli E, Estève J, Van Kappel AL, van Staveren WA. Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. Stat Med 1994;13:127-42
3. Van 't Veer P, Kardinaal AFM, Bausch-Goldbohm RA, Kok FJ. Biochemical markers for validation. Europ J Clin Nutr 1993;47:s58-s63
4. Willett W. Nutritional Epidemiology. Oxford: Oxford University Press, 1990
5. Kaaks R. Biochemical markers as an additional measurement in studies on the accuracy of dietary questionnaire measurements. Am J Clin Nutr (in press)
6. Loehnlin JC. Latent variable models. An introduction to factor, path, and structural analysis. Lawrence Erlbaum Associates, 1987
7. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol 1992; 3: 783-91
8. Dunn G. Design and analysis of reliability studies: the statistical evaluation of measurement errors. New York: Oxford University Press, 1989
9. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. Amer Stat 1983; 37: 36-48
10. GLIM. Oxford: Numerical Algorithms Group ltd, 1985
11. Jöreskog KG, Sörbom D. LISREL 7 a guide to the program and applicators Chicago: SPSS, 1988
12. SAS Institute Inc. SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1. Cary: SAS Institute Inc., 1989
13. Bentler PM. EQS: Structural Equations manual. BMDP Statistical Software Inc. Los Angeles, 1994
14. Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models. Part II. Stat Med 1993;12:937-48

Appendix 1. *GLIM program for BOOTSTRAP estimates of validity coefficients*

explanation and instructions for use

Input for this program is a dataset with observed values for questionnaire (Q), reference (R), and biochemical marker (M) measurements. The program creates an output dataset with 200 bootstrap estimates for validity coefficients of Q, R, and M (named rhoQ, rhoR, rhoM). If for a certain bootstrap sample no validity coefficients can be estimated because of one or three negative sample correlations, the validity coefficients are given the value 10 and the variable HWNEG is given the value 1 (otherwise 0).

to run the program

change the N value (now 61) which is the number of observations

and EXAMPLE.DAT which is the input file name

and EXAMPLE.RES which is the result file name

```
$MACRO CORR
$c macro for the calculation of correlation coefficients
$cal x=%1 : y=%2
$number mx=0 : my=0
$tab the x mean into mx
$tab the y mean into my
$cal x=x-mx : y=y-my
$cal %p=%cu(x*y)/%sqrt(%cu(x*x)*%cu(y*y))
$ENDMAC

$MACRO ONEOBS
$c macro for the drawing of one random observation from dataset
$cal var2=rand2(j)
$sseed var2
$cal alea=%sr(0)
$sort xQ,xR,xM,alea Q,R,M,alea
$cal yQ(j)=xQ(1) : yR(j)=xR(1) : yM(j)=xM(1)
$cal count1=(j<n)
$cal j=j+1
$ENDMAC

$MACRO ONESAMPLE
$c macro for the drawing of one bootstrap sample
$number var1=1 : var2=1 : j=1 : count1=1 : rQR=1 : rMQ=1 :
rRM=1 : rhoQ=1 : rhoR=1 : rhoM=1 : hwneg=0
$cal var1=rand1(i)
$sseed var1
$cal rand2=%sr(2047)
$while count1 ONEOBS
$use CORR yQ yR
```

```
$cal rQR=%p
$use CORR yQ yM
$cal rMQ=%p
$use CORR yR yM
$cal rRM=%p
$cal rhoQ = %sqrt( %if(rQR*rMQ/rRM<0 , 100 , rQR*rMQ/rRM ))
$cal rhoR = %sqrt( %if(rQR*rRM/rMQ<0 , 100 , rQR*rRM/rMQ ))
$cal rhoM = %sqrt( %if(rRM*rMQ/rQR<0 , 100 , rRM*rMQ/rQR ))
$cal hwneg = %if(rhoQ==10 ? rhoR==10 ? rhoM==10, 1,0)
$output 'example.res'
$print rhoQ rhoR rhoM hwneg
$output %poc
$cal count2= (i<200)
$cal i=i+1
$ENDMAC

$number n=61 : i=1 : count2=1
$unit n
$data R Q M
$dinp 'example.dat'
$variate 200 rand1
$sseed 76
$cal rand1=%sr(2047)

$while count2 ONESAMPLE
$stop
```

Chapter 8

General discussion

In the previous chapters, the central theme was measurement error in dietary intake data. More specifically, the difficulties in the assessment of vegetables, fruits, and antioxidant (pro)vitamin intake and, consequently, their relation to cancer risk was the focal point throughout the thesis. In this chapter, developments in dietary assessment from the first ecological studies on diet and cancer until the recent multicenter cohort studies will first be described. The concept of measurement error and its consequences for observed results, the assessment of dietary intake, and the estimation of measurement error will then be discussed in the context of (multicenter) cohort studies. Although many topics relate to dietary intake in general, examples concerning and specifics on vegetables, fruits, and antioxidant (pro)vitamins will be mentioned if applicable.

Dietary assessment in cancer epidemiology

During the 1970s, important hypotheses on diet and cancer have been derived from ecological studies using FAO per capita disappearance data and national cancer incidence or mortality rates. A well known example is the correlation between fat intake and breast cancer incidence [1]. Vegetables and fruits were not associated to cancer in these studies (see overview chapter 2). It has however been recognized that this type of studies was unsuitable for drawing causal inferences, among others because of unadjusted confounding [2,3].

Research in cancer epidemiology, therefore, shifted towards studies in which individuals were the units of observation. The majority of these studies were case-control studies, and most evidence for an inverse relationship between vegetable and fruit intake and cancer risk comes from them [4]. The development of dietary assessment methods to assess habitual diet before the onset of the disease received particular attention at this time, since the cases might have altered their diet because of the disease. However, differential bias in dietary intake by disease status could still not be excluded because current diet strongly influences recall of past diet [5,6].

Cohort studies are stronger by design than case-control studies because the exposure measurement precedes the (diagnosis of the) disease. However, in practice, problems also occur in the interpretation of results of cohort studies. The main reason for this is the combination of studying effects which are probably weak, small contrasts in dietary intake within populations, and relatively large measurement error in dietary

data [7]. This is particularly so in the case of vegetable, fruit, and antioxidant (pro)vitamin intake with respect to the occurrence of cancer.

In chapter 3, the intake of vegetables, fruits, and antioxidant (pro)vitamins was studied in relation to lung cancer incidence in a cohort of about 560 men in the Zutphen Study which was assembled in 1960. Dietary intake was assessed with the dietary history method. However, vegetable and fruit consumption was only assessed crudely at an aggregate level. This probably influenced the quality of the information negatively, but information on the validity of the data was not available to quantify the extent of the problem. Further, intake of vegetables and fruit was inversely related to smoking which implied that residual confounding could not be ruled out. This problem could not be studied with stratified analyses according to smoking status, because of the relatively small size of the study. All these limitations made a correct interpretation of the (intriguing) results difficult.

Some cohort studies tried to resolve part of the problems by increasing the size of the cohort. Examples of studies that included over 100,000 subjects are the Japanese cohort started by Hirayama in 1965 [8], The Nurses Health Cohort Study [9], the Cancer Prevention Study II [10], and the Netherlands Cohort Study [11]. The large scale resulted in the use of low-cost dietary assessment methods, of which some were very short [8,10]. As a consequence, the gain in power by the scale enlargement was counteracted by additional measurement error in the dietary data. Therefore, since the end of the 1970s, the improvement of low-cost methods that assess diet comprehensively, and the estimation of and correction for measurement errors in dietary data was further developed [12]. Parallel to this, the use of biochemical markers of intake was studied as alternative [13]. The studies described in chapters 4-7 of this thesis are examples of these developments.

However, soon it was realized that all low-cost dietary assessment methods are bound to include a considerable amount of measurement error, that the estimation of the size and structure of this error proved to be difficult [14] and that the number of suitable biochemical indicators of long-term intake is still very limited [15]. Therefore, to avoid the combination of large measurement error and homogeneous food patterns, ecological type of analyses recently returned on the stage. Prentice and Sheppard [7] suggested that between-population analyses are useful to estimate the size of effects if contrasts in dietary intake between populations are large, as is often the case, and suitable data on dietary and confounding factors are available for each population.

In chapter 2, we tested whether cohort average intake of antioxidant (pro)vitamins was related to cohort mortality of lung, stomach, and colorectal cancer in the Seven Countries Study. Although started around 1960, this study has important advantages compared to the traditional type of ecological studies. One major advantage is that exposure and outcome data apply to the same population as the study is a multicenter prospective cohort study. Further, nutrient intakes were analyzed in food-equivalent composites reflecting intake of the cohorts. Measurement error caused by the use of less reliable disappearance data or food composition tables could thus be avoided. However, important limitations are also associated with the study. Firstly, the number of observations is only 16 which makes multivariate analyses virtually impossible. Secondly, the populations are selected because of their differences in culture. The differences in dietary intake may go together with differences in other factors, which might confound the relationships. And thirdly, dietary data may have been differentially biased across the 16 cohorts because the food records were not highly standardized across countries, i.e. some were weighed records, either by the participant or by outside investigators, and others estimated records. An additional source of bias may be the small size of the subsamples of cohort members that kept food records. Because the impact of these limitations on the observed results is unknown, their meaning remains unclear.

At the end of the 1980s, the European Prospective Investigation into Cancer and Nutrition (EPIC) was planned, the first multicenter cohort study on diet and cancer with all exposure, confounder, and outcome data to be collected at individual level [16]. Data analyses of this type of studies can be performed both within and between cohorts. In the analyses at ecological level, adjustment for confounding variables can be done at individual level or stratified analyses may be conducted. However, dietary assessment in multicenter cohort studies introduced a new combination of requirements. Apart from the requirements for large-scale cohort studies, the method had to be suitable for the research populations in all centers. As this seems extremely difficult in practice, an alternative approach was chosen in the EPIC-study. The approach involves the use of country or center-specific dietary assessment methods in conjunction with a calibration study in subsamples of the cohorts to correct for differences in systematic measurement error across cohorts [17,18].

Measurement error in dietary intake data, and consequences for (multicenter) cohort studies

Many dietary assessment methods are available, which each have their merits and limitations in specific situations and none is error-free [19-21]. In order to clarify different types of error we will first introduce a measurement error model, according to Kaaks *et al.* [22]. In this model, measured intake includes both systematic and random error. It is assumed that systematic error is linearly related to true intake, that true intake is normally distributed, and that the random measurement error is independent of true intake, and normally distributed with mean 0 and variance σ_ε^2.

$$M_i = \alpha + \beta T_i + \varepsilon_i$$

with M_i= measured intake of individual i

T_i = true intake of individual i

α = systematic constant measurement error

$(\beta-1)T_i$ = systematic proportional measurement error of individual i

β = proportional scaling factor

ε_i = random measurement error of individual i

The systematic constant error term α indicates the average tendency to over- or underestimate intake by a constant amount. The proportional scaling factor β reflects the average tendency of individuals to overestimate ($\beta>1$) or underestimate ($\beta<1$) intake by an amount which is proportional to the level of true intake. The so-called 'flat slope syndrome', referring to the overestimation at low and underestimation at high levels of intake [23] ($\beta<1$), is an example of this type of error. The total random error - random refers here to its effect at group level - can be divided in two components. The first is within-subject random error, and the second is the unpredictable part of the individual's bias, also called the random bias or subject-specific bias. Hence, the latter type of error is random at group level, but systematic at individual level. Within-subject random error will not be reproduced when repeated measurements are taken, whereas random bias will be reproduced. At population level, the systematic constant error, the proportional scaling factor, and the size of the variance of the random error relative to the variance of the true intake are the indicators to describe the different dimensions of measurement error. The validity coefficient, i.e. the Pearson correlation between measured and true intake is a measure which incorporates both the latter variances [24]. Intuitively, it may be clear that the measurement error parameters are

determined by the particular dietary assessment method and the population in which it is applied [14].

If methods that assess actual intake are used to estimate habitual intake, the selection and number of days covered introduce measurement error in addition to the error made in the dietary intake assessment per se [14]. For example, if assessment is done within one season or on a selection of specific days of the week this may introduce bias, i.e. systematic measurement error if it concerns all subjects, or random bias if it is equally distributed among the study population. The random variation in daily intake forms part of the within-subject random error and can consequently be reduced by taking more measurements.

It is important to realize that the presented model includes several assumptions which may not be true. It could for example well be that systematic error is not linearly related to true intake but in another way, or that random measurement error is not normally distributed. Further, it is well possible that the bias is related to other factors. This type of error is called differential bias. One could easily imagine, that (part of) what is considered random bias is in reality differential bias, i.e. explained by other factors not in the measurement error model. Very little is known about differential bias, although one factor for which differential bias in energy intake is established (in Western populations) is body mass index [25]. Another situation in which differential bias seems likely, occurs when the individuals that provide information about food consumption belong to several culturally diverging populations. Population differences in for example educational level, social desirability, and attitudes towards diet might then introduce differential bias. In multicenter cohort studies special attention should be given to this possibility.

During the last decade the effects of measurement error in exposure on measures of association in epidemiological studies and on power and sample size have been worked out theoretically for relatively simple situations [24]. The consequences of measurement error on measures of association will be illustrated below for some situations relevant to (multicenter) cohort studies. Further methodological research is needed to allow for measurement error in more complex circumstances, for instance correlated errors in exposure and confounding variables.

In within-cohort analysis, dietary intake is often considered in quantiles. Random measurement error is the only type of error that effects measures of association for these analyses. In case no assumptions are made about the distribution of intake and

of measurement error the effects of misclassification on odds ratios may be calculated as was done in the simulation study in chapter 4. In this approach the misclassification matrix, i.e. the matrix of the proportions (p_{ij}) of those with true exposure category j who will be classified into category i, relates the true distribution of intake in a population to the misclassified distribution [24]. The resulting observable distribution of exposure among cases and controls can then be used to calculate the observable odds ratios for comparisons of disease risk for each category of exposure versus the reference category. One example in chapter 4 concerned total carotenoid concentrations in EDTA-plasma after 2 years of storage at -20 °C. Using the misclassification matrix from the stability study, it was simulated that true odds ratios of 0.64 and 0.40 for the second and third versus the first tertile of carotenoid level would have been observed as 1.10 and 0.83 due to the error in the concentrations of the stored samples. In the case of two categories of exposure random measurement error will result in attenuation of the odds ratio. When there are more than two categories of exposure, general conclusions cannot be drawn about the effect of non-differential misclassification on the odds ratio for each category [26].

If the joint distribution of true and measured exposure can be assumed to be bivariate normal then the misclassification matrix for a given validity coefficient can be computed. Using this matrix and the true distribution of cases and controls over quantiles of intake, the expected observed relative risk can be calculated. This is the approach adopted by Walker and Blettner [27] and Marr and Heady [28]. De Klerk *et al.* [29] provide values for expected measured relative risks comparing extreme quartiles, for given true relative risks and validity coefficients. For example, in chapter 6 the relative validity among men of vitamin C intake assessed with the FFQ was 0.45. If this correlation was a valid estimate of the validity coefficient, a true relative risk of 2.00 comparing subjects in the lowest to subjects in the highest quartile of intake is then expected to be observed as 1.31. If the reproducibility coefficient of 0.75 was taken as a valid estimate of the validity coefficient, a much higher relative risk of 1.59 would have been expected to be observed. Clearly not all random measurement error is then accounted for.

When the strength of associations within-cohorts has to be compared or combined, as is the case in multicenter cohort studies, dietary intake could better be considered in a scaled quantitative way instead of quantiles [30]. For this type of analyses, in which relative risks are calculated for a given increment in intake, both

random and systematic proportional measurement error may influence measures of association. Under the assumptions that true intake and random measurement error are normally distributed, that the association between true intake and the disease is log-linear, and that the overall incidence of diet-related disease is low, then the log relative risk will be biased with a factor which is equal to the square of the validity coefficient divided by the proportional scaling factor [17]. Coming back to the example of the validity of vitamin C intake in men from chapter 6 (with an estimated proportional scaling factor of 0.32), this would mean that the bias factor would be $0.45^2/0.32 = 0.63$. A relative risk of 0.5 for 10 grams increment in intake would then be observed as 0.65. An alternative way to estimate the bias factor is by the linear approximation method described by Rosner *et al.* [31], in which true intake is regressed on measured intake. The bias factor is equal to the slope of this regression.

In case of differential bias the error is related to another factor of interest. Differential bias introduces a spurious association between the factor of interest and the measurement of the dietary factor. If this factor is also related to outcome, differential bias may result in an over- or underestimation of the crude measure of association between the exposure and outcome variable [32].

For between-cohort analysis, groups of individuals are the unit of analysis. Random measurement error does not play a role in this situation, as its expectation at group level is zero. Systematic measurement error can be expected to differ across cohorts. Subtracting the systematic bias for average intake from average intake of each cohort will adjust for these differences in error across populations [17].

If compatible with one another, within and between cohort evidence may be combined. The within and between cohort estimates of the measures of association may be weighted according to their imprecision (including that due to measurement error), to obtain an overall estimate, as described by Kaaks *et al.* [17]. The chance that not all within and between cohort evidence is compatible with each other seems larger in multicenter studies in which the centers are very different from one another. The Seven Countries Study with centers as different as Northern Europe, Southern Europe and Japan is an example where this problem may arise.

Assessment of usual dietary intake

For large scale (multicenter) cohort studies the comprehensive assessment of dietary intake, rather than the estimation of a few nutrients, foods, or food groups is of

importance. The reason for this is that many hypotheses usually need to be tested, and that many dietary variables including total energy intake may play a role as confounding variables. Further, estimating long-term dietary intake (period of years/decades/lifetime) is the conceptually relevant exposure in cancer epidemiology. Together with the limitation of resources in relation to the number of subjects, this often results in the choice of an extensive self-administered semi-quantitative food frequency questionnaire (further denoted FFQ) as dietary assessment method [33]. As a considerable amount of measurement error cannot be avoided with FFQ, a loss of power will be the result, even if the measurement error is estimated and corrected for. For this reason a second assessment of dietary intake may be desirable for all subjects [24]. After discussing dietary assessment with FFQ, different options of such a second measurement will be given.

A FFQ estimates how frequently certain specified foods are eaten. The period of reference is sometimes undefined as 'usually' or 'habitually', or it is specified most commonly ranging from the past month to the past year. Information on usual portion size may be asked or a standard portion size may be given for which the consumption frequency is asked (see also later). Commonly most of the answer options are in closed form. Other names for similar types of questionnaires are: diet history questionnaire [34], food use questionnaire [35], quantitative food frequency questionnaire [36], or frequency and amount questionnaire [37].

The development of such a FFQ should be carefully done, and pretesting of (different versions of) the questionnaire is extremely important [37]. A FFQ should be focused on the meal pattern, educational level, and use of language of the population in which it is to be used. The same FFQ should therefore not be used to estimate usual intake of populations with different dietary patterns, which limits its use in multicenter cohort studies. Nor should the same FFQ be used at different points in time for a population that changed diet largely during the interval period. For the selection of food items for which consumption frequency is asked a data-based approach as described by Block *et al.* [38] should preferably be used. This method, in which items are selected on the basis of their contribution to intake at group level using an external dataset, was also applied for the development of the Dutch EPIC FFQ as described in chapter 5. An improvement of this approach involves the selection of (additional) items based on the amount of explained variation in intake, as was already done by Heady

[39]. Non-dietary sources of nutrients, like supplements for vitamins, may also be important to include in the questionnaire [40].

The Dutch EPIC FFQ incorporated the following aspects in an attempt to improve the questionnaire. Few data are, however, available in which the effects of this types of design aspects are quantified. The gains in validity are likely to be quite modest. Nevertheless, even modest additional increments are worth trying to achieve.

1. The questionnaire started with a question on usual meal pattern. This was done as a memory aid to the subject for the estimation of usual consumption frequencies, and is used in the stages of data checking and processing. For example, large discrepancies between the number of dinners and the number of times rice, pasta, potatoes or other staple foods are eaten are checked with the subject.
2. Most food items were ordered according to the usual daily meal pattern in the Netherlands, with the object to make remembering average intake easier. Among a Latin population, Boutron *et al.* [41] observed better relative validity for a diet history questionnaire with foods ordered by meal than for a questionnaire in which the same foods were ordered by food group.
3. The consumption frequencies for global categories of foods are asked first in an accurate way, followed by questions on the consumption of specific foods with more crude answer options. Asking global categories first, may help the respondent to prevent the tendency to overestimate consumption frequencies of single food items, and may give the researcher a tool to correct the individual frequencies [42,43]
4. Brand names are asked for margarines used on bread and for cooking fat because of their heterogeneous composition. However, in a study of Willett *et al.* [44], the ignorance of responses to open-ended questions on cooking oil, cold cereal type, and multiple vitamin brand in a FFQ effected only the relative validity of some of the micronutrients as compared to diet records 4 years earlier.
5. Questions on portion sizes, referring to food photographs, were added for 28 food items. Information on the range of portion size in an external dataset was used to decide for which products portion size varied widely, and determine which amounts could best be photographed to allow for a sufficient range of expression of portion size in the answer options. The need to ask for portion

size has been debated, as was discussed in chapter 5. Our own validity study as well as those of others indicated that in general little validity in the ranking of subjects was gained by using food photographs for the estimation of portion size. For instance, the Spearman correlation between vegetable intake estimated by the Dutch EPIC FFQ and with 12 24-h recalls changed from 0.38 to 0.36 for men and from 0.31 to 0.25 for women when portion sizes assessed by food photographs were replaced by standard portion sizes. For estimating median intake at group level, quite some improvement was observed for most food groups, although this was not the case for vegetable intake among men (chapter 5). Further food photographs, due to their attractiveness, may have a positive influence on response and motivation for filling out the questionnaire.

6. We created the possibility of feedback to the EPIC-participant in case of important missing or inconsistent data. The costs of this may be limited, and the procedure may be standardized by optical reading of filled out FFQ in combination with software for an automated check on missing, inconsistent, or unlikely data. If this is done when the participant visits a study center (e.g. for anthropometric measurements, drawing blood) any queries can immediately be clarified.

To calculate the intake of nutrients, information is required on the composition of each food in the questionnaire. The use of food composition tables will invariably introduce measurement error. This is due to the assumption that the nutrient content of a specific food is approximately constant, non-random sampling of foods for the chemical analyses, errors in the chemical analyses themselves, incorrect use of values from other food composition tables, and missing data [45]. For antioxidant (pro)vitamins the errors are probably larger than for macronutrients, because the variation in the vitamin content of foods is generally much greater than the variation in macronutrient content [45], and because still many values in food composition tables are determined with less reliable analytical techniques than those currently available [46].

For some diseases it may be argued that not the intake of the nutrient per se, but the amount that is available for utilization by the body, i.e. its bioavailability, is of major importance. Bioavailability is the result of the interaction between the nutrient in question with other components of the diet, either from the same food or from other

foods, and with the individual consuming the diet [45]. For ß-carotene large differences in bioavailability exist according to food source [47,48] and a correction factor for bioavailability applied to the nutrient content may possibly improve data on ß-carotene available to the body. One would for example expect a better correlation between ß-carotene intake and blood concentrations if the intake is adjusted for bioavailability. As in the Netherlands the major sources of ß-carotene are yellow/orange and green vegetables, with a low bioavailability, this might partly explain the observed very low correlations between intake and serum concentrations of ß-carotene in chapter 6.

It is particularly difficult to describe the validity of FFQ in general because of large differences between the questionnaires and the populations in which they are applied. Correlation coefficients with data from food records or diet recalls vary widely, but usually range from 0.3 to 0.7 for most nutrients and food groups. No questionnaire is available with all correlation coefficients for a comprehensive list of nutrients and food groups above 0.6 [49]. The relative validity for the ranking of individuals according to vegetable intake is generally poor, and that of fruit intake moderate (see chapter 5). As a consequence low relative validity is also often observed for antioxidant (pro)vitamins. The Dutch EPIC FFQ also performed poor in the ranking of the three antioxidant (pro)vitamins and vegetables. It is questionable whether a better relative validity for vegetable intake can be obtained with a food frequency approach: asking an average consumption frequency and eventually average portion size may not be the correct way to assess intake of infrequently eaten individual vegetables. An alternative approach will be proposed later. The population mean intake and other distribution characteristics may be estimated with varying validity using FFQ. The tendency to overestimate individual food frequencies is counteracted by underestimation due to an incomplete list of foods included in the questionnaire; the final balance being questionnaire specific.

Few surveys have studied the error structure of FFQ. Beaton [50] deduced (under certain assumptions) from several published validity studies, that FFQ appear to loose some of the real variance between subjects and at the same time introduce new (spurious) between-subject variance. He concluded that the loss of variance might have arisen because it is impossible to ask about all foods consumed and hence that the reductionism and summation needed in responding to the questionnaire necessarily resulted in some underreporting of variation. At the same time he postulated that the new variance added had arisen from instrument errors that were systematic as far as

the particular subject was concerned but that differed between subjects. Flegal *et al.* [51] partitioned differences in macronutrient intake estimates between a FFQ and diet records into several components. It was shown that frequency differences were the main source of poor agreement in relative ranking, serving size differences contributed mainly to disagreement in group mean intake, and differences in nutrient composition had only minor effects.

Second dietary measurements

A second assessment of dietary intake for all subjects in the cohort study may be very useful to compensate for the loss of power because of unavoidable measurement error [24] associated with FFQ. If within-subject random error in FFQ would have been large, a considerable gain could be achieved by a repeat of the same questionnaire. However, in general, reproducibility of FFQ is quite good with median correlation coefficients around 0.7. This was also the case for vegetable, fruit and antioxidant (pro)vitamin intake assessed by the Dutch EPIC FFQ (chapters 5 and 6). Therefore, possibly more may be gained with an additional measurement of reasonable validity with different, preferably independent, types of error. In this way part of the random bias may also be reduced. A combination of FFQ with biochemical indicators of intake, or with a method of actual intake may therefore be more efficient in practice than repeated measures of the same questionnaire. In the analysis phase both types of measurements may be combined by calculating a weighted average value for each person, with the proportion of true intake explained by the methods as weights. For this reason validity coefficients need to be estimated for both methods.

Before discussing in more detail the use of dietary methods of actual intake and biomarkers, a remark will be made about the use of repeated measurements for another purpose, i.e. capturing changes in dietary intake over time. For this purpose the same dietary assessment method should be used at the repeated occasions, because otherwise changes in intake cannot be distinguished from differences due to the methods. However, the method of choice should not be a FFQ which is based on a predefined lists of foods, but biomarkers or a method with open structure, such as the diet history, diet recall, or diet record. In chapter 3, diet was assessed three times during a 10 year period using the same dietary history method. The relative risks for a high versus low intake were much stronger when subjects with stable intake were compared instead of average intake of all subjects. For studies with a long follow-up period it is important

to know whether changes in intake in the follow-up period differed across individual (within-cohort analyses) or across cohorts (between-cohort analyses). Such changes might confound the association between intake at the start of the follow-up period and outcome.

Dietary methods to assess actual intake have been described and discussed extensively in textbooks [52,53]. In the context of large-size cohort studies a self-administered method which can be explained in writing would be most appropriate for financial reasons. A suggestion to assess actual intake of total and individual vegetable consumption would be to ask subjects to record vegetable consumption (type and amount) for one week in a diary combined with a list of vegetables for which the last time it was consumed is asked. It would be interesting to validate this approach.

Up to date, for many nutrients, non-nutritive substances, foods, and food groups no suitable biochemical indicator of long-term intake is available. This limits the use of biomarkers as a second comprehensive method to assess dietary intake. For those biomarkers that are available, proper use is not straightforward and it often requires knowledge of complex metabolic and pathologic pathways [13]. For large-size studies, the material in which biochemical indicators are determined is usually restricted to easily accessible tissues. Most commonly blood is used, but urine, hair, nails, saliva, or fat biopsies are also good alternatives. Many of the biochemical indicators in urine require 24-h collections, which are a large burden to the subject and the researcher, and are therefore not often used in large-size studies.

In order to limit costs, samples are often first stored after collection and analyzed afterwards for those subjects who actually develop cancer and those from a subcohort or matched healthy controls. This means that the concentrations of the indicator in the sample must not be affected by storage conditions. In chapter 4 it was shown that for blood carotenoids and vitamin E storage at -20 °C is not appropriate for prospective studies, and this is also the case for vitamin C [54]. Storage below -70/-80 °C seems therefore needed [54,55], and if possible even colder storage is advised, e.g. storage in liquid nitrogen at -196 °C as is done in the EPIC-study. Every study in which biological material is stored should have an in-built check for the chemical stability under storage conditions, i.e. chemical analyses of interest should be done immediately after sample collection on a small subsample of the cohort, and the same tests repeated over the following years [13]. Other factors important to any discussion of quality control in clinical chemistry, also apply to the quality of biochemical

indicators of dietary intake. It is therefore also important to define optimum collection and handling conditions and analytical techniques.

Recently, potential biochemical indicators reflecting intake of fruit and vegetables have been investigated. Campbell *et al.* [56] studied the use of 5 different carotenoids, as these compounds are widely distributed in vegetables and fruits with little contribution from other sources. The sum of lutein, ß-cryptoxanthin, α- and ß-carotene showed a correlation of 0.54 with total intake of vegetables and fruits, although α-carotene concentration alone correlated just as well. These results may however be population specific depending on the type of vegetables and fruits eaten.

The most commonly used biochemical marker for ß-carotene intake is the plasma or serum concentration, which reflects intake within a period of a few days or weeks. Plasma and serum levels are influenced by the rate of intestinal absorption, the efficiency of the enzymatic transformation into retinoids and the rate of clearance from plasma. Between-subject variation in the plasma response was found to be substantial in a supplementation study [54]. In chapter 7, the validity coefficient for serum ß-carotene levels as a measure of habitual intake was estimated to be 0.32 for Dutch males on the basis of a triangular comparison with intake obtained by FFQ and 24-h recalls. Recently, the use of adipose levels of ß-carotene was evaluated. Weak correlation coefficients of 0.30 for men and 0.12 for women were observed between adipose tissue levels and intake assessed by FFQ. Whereas correlations found between plasma levels and the FFQ were in the same order of magnitude (-0.07 and 0.33 for men and women, respectively). Within-person variability of adipose tissue levels was higher than of plasma levels [57].

Vitamin C exhibits a strong correlation between recent intake and biochemical indices, but with the exception of levels in leukocytes, do not adequately reflect usual individual intake. Plasma levels show a characteristic S-shaped curve with the steepest change in plasma levels between about 30 and 90 mg intake per day for adults. Amounts in the buffy coat do not exhibit the lower threshold effects that plasma levels do, and hence provide a more sensitive measure of lower intakes. Urine vitamin C is a potential marker for high intake, because above about 60-70 mg/day excess of the circulating vitamin is being excreted in the urine. The measurements of vitamin C presents certain problems because of its instability. An acid to stabilize ascorbate has to be added in the tube before, or closely after taking the sample [54]. Based on a triangular comparison with vitamin C intake obtained by serum concentrations, FFQ

and 24-h recalls in German males, the validity coefficient for the serum levels was estimated to be 0.39 in chapter 7.

Vitamin E naturally occurs in the form of four tocopherols and four tocotrienols, of which α-tocopherol has the highest biological activity. Plasma levels of vitamin E depend on the dietary intake of the vitamin and are related to the amount and type of lipoproteins and other plasma lipids [54]. Although within the same subject there is good correspondence between dietary intake and plasma levels of vitamin E, the relationship is less straightforward when comparing subjects. Adjustment for cholesterol and triglyceride levels usually improves the correlation between α-tocopherol levels in plasma and diet, although a partial correlation coefficient adjusted also for age, sex and total energy was only of the order of 0.3 in a methodological study [58]. The weak correlation is partly because the absorption is incomplete and variable (between 20 and 80 percent in published studies) and partly because the extent of absorption declines with increasing amounts per meal [54]. Applying the method of triads described in chapter 7 to the vitamin E data of the Dutch EPIC validity study, yielded an estimated validity coefficient of 0.32 (unpublished results). Schäfer and Overvad [59] reported that adipose tissue vitamin E is strongly associated with intake assessed with the dietary history method (r=0.76), whereas Kardinaal *et al.* [57] observed much lower correlations with a FFQ of 0.16 for men and 0.30 for women. The difference might be explained by the fact that the subjects of the former study included vitamin E supplement users which increased the range of intake.

For vegetables, fruits, and the three antioxidant (pro)vitamins, there is no biochemical indicator available that indicates the level of intake; they are only suitable for ranking subjects. Also for other nutrients, only few biochemical indicators exist that may be used to estimate the magnitude of intake. One example of this is 24-h urinary nitrogen excretion for protein intake [60,61].

Estimating measurement error in data on usual dietary intake

Until recently the estimation of measurement error in dietary data was mostly done by comparing the data with those obtained from a method with known better validity. The measure of relative validity thus obtained was mainly used qualitatively to decide whether a dietary assessment method was of adequate validity for use and to aid in the improvement of the dietary assessment method. Repeatability studies in which within-

subject random measurement error is estimated are also traditionally carried out. During the last decades, the focus of studies on measurement error has shifted towards more quantitative and statistical aspects. Important additional purposes of estimating measurement error are the calculation of sample size requirements or statistical power for studies in which the dietary method is (to be) used, and the correction for effects of measurement error in later statistical analyses of these studies. Sufficient evidence is now available that this adjustment is a must for within-cohort analyses [49,62].

Studies in which the total (relevant) measurement error is estimated are called validity studies [24]. In order to decide how validity can best be estimated, the theoretical conceptualization of what a method has to measure needs to be specified. This specification should be in terms of time-frame (e.g. intake over a month, year, habitual intake); the foods, food groups, nutrients, or other substances of interest; the population in which it is to be applied; whether individual or group level information is of importance, and how the dietary data will be used in the analyses (for example in quantiles, or continuously).

Other studies which aim to estimate measurement error of some kind also exist, like for instance the already mentioned repeatability studies. A second example are stability studies, such as the study described in chapter 4, in which measurement error in blood (pro)vitamin concentrations due to storage at -20 °C was estimated. If the concentrations of the (pro)vitamins have the purpose of reflecting dietary intake, this measurement error is obviously only part of the total measurement error. The total error is also determined by how well the true levels of the (pro)vitamins in the body reflect intake, and by factors in obtaining the material and preparing it for storage [54].

Recently, the term calibration study has been introduced in nutritional epidemiology [17,18]. It is used for studies in which (only) those parameters are estimated which are needed to correct measures of association for given increments in intake for measurement error. Calibration is suitable for both within-cohort and between-cohort analyses [17]. The difference with validity studies lies in the fact that not all parameters of the measurement error model are estimated separately. More concretely, the systematic measurement error and - for within-cohort analyses - the variance of the calibrated measurements are estimated, but not the separate variances of the random measurement error and true intake. A calibration study can consequently not be used to estimate the consequences of measurement error on relative risks comparing quantiles of intake. Although parameters estimated from a validity study

may also be used to correct relative risks for given increments in intake, the conduct of a validity study is less efficient for this purpose than a calibration study [63].

In both validity and calibration studies, the test method needs to be compared with a method of reference. The choice of the reference method and statistical analyses for different aims will be discussed. A short remark will also be made about the sample size of validity and calibration studies. Considerations on other aspects of design of these studies can be found in several reviews [e.g. 23,33,49,64].

For the estimation of group level systematic measurement error the nutrient or food intakes assessed with the reference method should be unbiased at group level. The estimations can be obtained from regressing intake as assessed with the test method on intake as assessed with the reference method. The intercept and regression coefficient thus obtained are estimates of the systematic constant error (α) and the proportional scaling factor (β), respectively [17].

However, few dietary methods exist which give unbiased estimates of group level intake. In populations which are in equilibrium, these are 24-h urinary nitrogen excretion as a measure of protein intake [60], and energy expenditure assessed with the doubly labeled water method as a measure of energy intake [65]. Both methods measure actual rather than habitual intake and the measurements therefore need to be taken with avoidance of day-of-the-week, seasonal or any other systematic effects at group level. Further, systematic effects due to practical issues, like incomplete urine collections need to be avoided.

Weighed dietary records, cross-check dietary history interviews, and carefully employed 24-h recalls are also used as reference methods for the estimation of systematic measurement error but this may be inappropriate as data obtained with these methods probably include systematic error. However, for most populations these methods may be less biased than other methods and will be the only option to estimate crudely systematic measurement error for an extensive list of nutrients and foods or food groups. For instance for intake of vegetables, fruits, and antioxidant (pro)vitamins, no better alternative seems present. In chapters five and six, intakes of nutrients and food groups assessed by the Dutch EPIC FFQ were compared at group level with those based on the average of twelve repeated 24-h recalls. For protein intake an additional comparison was made with intake based on the average of four 24-h urinary nitrogen excretions. The conclusion of the latter comparison (underestimation by FFQ) was not

in accordance with that of the first comparison (overestimation by FFQ), indicating systematic underestimation in the 24-h recall data.

For the assessment of the variance of the random error and of true intake and the validity coefficient (applicable to validity studies) two options are in theory possible. These are:

1. to use a reference method, which does not produce random measurement error itself;
2. to use two or more reference methods, which are each linearly related to true intake and whose random measurement errors are mutually independent of each other and of the random measurement errors of the test method.

In the case of a single reference method without random measurement error, the variance of true intake is equal to the variance of the reference measurements, while the variance of the random error in the test method is equal to the difference between the variances of the reference and test methods. The correlation between the reference and the test measurements provides the validity coefficient [22].

However, when habitual intake of free-living individuals is of interest no reference method exists which is free of random measurement error [33]. The within-subject random error can be avoided as much as possible by taking sufficient repeated measurements or, in the case of normally distributed variables, may be estimated and corrected for by taking at least two repeated measurements [66]. In chapter 5, relative validity of specific vegetables was not reported, because the random measurement error in the mean of twelve repeated 24-h dietary recalls, was considered too high due to the large within-subject variation in intake of specific vegetables. Within-subject random error could not be estimated because of the non-normal distribution of specific vegetable intake. To a lesser extent this is also the case for specific fruits. The random bias cannot be eliminated or estimated by taking repeated measurements. In practice, two different situations are likely to occur, i.e. random measurement error of the reference method is uncorrelated or is positively correlated with that of the test method.

1a. If random measurement error in the reference method is independent of that of the test method, the random error of the test method will be overestimated, and the validity coefficient underestimated. Uncorrelated random error might be expected if the test method is a food consumption method and the reference method a biomarker [33]. The correlation coefficients between protein and

vitamin intakes estimated by FFQ and biomarkers in chapter 6 seem examples of underestimated validity coefficients.

1b. If random measurement error of the reference method is positively (but not perfectly) correlated with that of the test method, one cannot tell whether the random measurement error and validity coefficient of the test method will be under- or overestimated [22]. The result depends on the balance between the size of the independent random error in the reference method (gives underestimation of the validity coefficient) and the strength of the correlation between the random errors of both methods (gives overestimation of the validity coefficient). A positive correlation between random errors can be expected if the test and reference methods both rely on subject-obtained information about food consumption. Subject-specific tendencies to over- or underestimate are the reason for this. The positive correlation, and the likelihood of overestimating the validity coefficient, may be kept as low as feasible by choosing a reference method that is essentially different from the test method [33]. For a test method in which habitual intake is assessed the reference method should preferably assess actual intake [64]. The correlation coefficients between the values of the FFQ and 24-h recalls in chapters 5 and 6 may be either over- or underestimates of validity coefficients for intake assessed by FFQ.

If two reference methods are used which are each linearly related to true intake and whose random error are mutually independent of each other and of the random error in the test method, the validity coefficient may be estimated using the method of triads, as was shown in chapter 7. The variance of true intake may be derived from this by multiplying the variance of measured intake by the square of the validity coefficient of measured intake by either of the three methods. The variance of the random error in the test method is equal to the difference of the variances of measured intake by the test method and true intake.

The mutual independence of the measurement errors of two dietary assessment methods seems possible, but when mutual independence of three methods is required, positively correlated errors may be expected because two methods will usually rely on food consumption data. In chapter 7, we argued that the estimated validity coefficients for the FFQ using the triangular comparison with 24-h recalls and biomarker values, are probably overestimates because of positive correlation between random errors in

the FFQ and 24-h recalls. It was also discussed that a more appropriate conclusion was that the true validity coefficient lies in a range for which the minimum is the correlation between FFQ and biomarker values and the maximum is the validity coefficient obtained by the method of triads.

For within-cohort calibration the requirement of no random measurement error in the reference method is not necessary. What is needed is a single reference method without systematic proportional bias and with random errors uncorrelated to those of the reference method [17,18,31]. The linear approximation method described by Rosner *et al.* [31] may be used to perform the calibration. In this method, intake obtained with the reference method is regressed on intake obtained with the test method. The inverse of the regression coefficient (λ) thus obtained is used as a correction factor for the effect estimate relating one unit increment in intake to outcome in a log-linear way. The regression coefficient may also be used to adjust the confidence interval of the effect estimate. The variance of the calibrated measurements may be estimated as the square of the regression coefficient (λ) times the variance of the test method.

The two requirements for the reference method for within-cohort calibration were already discussed separately above. Few reference methods seem able to fulfill both, i.e. the biomarkers 24-h urinary nitrogen excretion to estimate protein intake [60] and the doubly labeled water technique to estimate energy intake [65]. When other nutrients or food groups are of interest (single) weighed dietary records and 24-h recalls may be the best approximations of these requirements in the case of a test method that asks for habitual intake.

All relevant parameters of validity studies may also be estimated using structural equation models. In these models the relations between measured and true intake are described, as well as the relation between intake measured in different ways. Parameter estimates of the measurement error model can be derived from fitting the theoretical first and second moments predicted by the structural equations model with the observed moments estimated from actual measurements [22,67]. The advantage of such a model is that other assumptions than those specified above may also be introduced in the model. For example, in a study with several methods on several spaced occasions, Plummer and Clayton [67] assumed that only different methods on different occasions do not have correlated errors.

Another issue which should be considered when choosing a reference method is its effect on response. Especially if measurement error is estimated for correction

purposes of the measures of association the validity or calibration study should be conducted in a representative sample of the main study. A high response to the reference method(s) is then of major importance to ensure the representativeness. Of the methods mentioned above a single 24-h recall seems to have the smallest burden to the subject, a cross-check dietary history takes more time and a weighed dietary record requires a considerably high cooperation from the subject [21]. In many populations, the collection of blood gives a moderate response rate only. Even lower rates can be expected for the collection of 24-h urine or the participation in a doubly labeled water protocol. Thus, the two methods which seem most appropriate as reference method, in the sense that they provide (nearly) unbiased estimates and have random errors uncorrelated to those of FFQ, are the least appropriate for their effects on response. For response reasons, it is also better to estimate random within-subject measurement error in the reference method based on two repeats and use this estimate to adjust the validity coefficients instead of trying to eliminate random within-subject error by taking many repeats.

The response of the validity study described in chapters 5 and 6 was quite low, and it had to be concluded that the estimated relative validity may be an overestimate of true relative validity because of a possible selection towards more health-conscious subjects. As the design of validity studies often includes several different dietary assessment methods administered repeatedly, it is in practice difficult to obtain a high response. A solution would be to use incomplete block designs in which many combinations of methods and repeats are administered, but only relatively few are used in any one subject [67]. Such a design would also prevent negative effects of intensively surveying individuals. For calibration studies, a high response might be obtained more easily as only one reference method is needed and repeated measurements are not necessary to reduce random within-subject error [63].

The number of subjects in pilot-phase validity studies is usually around 100. For estimates of the validity coefficient based on one reference method, confidence limits will be rather wide with this sample size, especially if subgroup analyses are carried out. For example a validity coefficient of 0.5 has a 95%-confidence limit of 0.33-0.63 in case of 100 subjects. Confidence intervals become even wider with lower correlations [49]. If such imprecise estimates are used to correct measures of association, confidence intervals of these measures will also become considerably wide. Larger sample sizes of several hundreds of subjects are therefore recommended. The

precision of estimates of systematic measurement error, of validity coefficients based on two reference methods, and of the variance of calibrated measurements depends both on the number of people and on the random error in the reference method. If random error in the reference method is large, more subjects are needed to obtain the same precision. In chapter 7 this was illustrated for validity coefficients estimated with the method of triads. Kaaks *et al.* [63] discuss the approximate sample size required to have a sufficient level of accuracy in dietary calibration studies. They conclude that the cost for calibration is more efficiently reduced by increasing the total sample size than by taking replicate reference methods.

Conclusion

Measurement error in dietary data is a true handicap for epidemiologic research, but cannot be avoided. Intake of vegetables, fruits, and antioxidant (pro)vitamins are examples of dietary exposure variables which are particularly prone to error. Random measurement error frustrates the tests whether an association does exist, and both random and systematic measurement error affects the assessment of the size of an association. Further progress lies in improvement and continuous adaptation of dietary assessment methods to specific aims, but probably even more in the understanding of the nature and magnitude of error in dietary data and analytical methods that recognize and cope with that error. Many of the models currently in use to estimate and correct for measurement error, are fed with assumptions which do not necessarily correspond with the true situation. A more appropriate approach might therefore be to allow for the uncertainty of the error structure and perform sensitivity analyses to test different assumptions both at the extreme end and for more likely situations. Further study is also needed on factors that determine systematic bias at the individual level, since it seems unlikely that this bias is truly random at group level.

References

1. Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int J Cancer 1975; 15: 617-31
2. Piantadosi S, Byar DP, Green SB. The ecological fallacy. Am J Epidemiol 1988; 127: 893-904
3. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. Int J Epidemiol 1989; 18: 269-74
4. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer. I. Epidemiology. Cancer, Causes Control 1991; 2: 325-57

5. Bakkum A, Bloemberg B, Van Staveren WA, et al. The relative validity of a retrospective estimate of food consumption based on a current dietary history and a food frequency list. Nutr Cancer 1988; 11: 41-53
6. Friedenreich CM, Slimani N, Riboli E. Measurement of past diet: review of previous and proposed methods. Epidemiologic Reviews 1992; 14: 177-96
7. Prentice RL, Sheppard L. Validity of international, time trend, and migrant studies of dietary factors of disease risk. Preventive Medicine 1989; 18: 167-79
8. Hirayama T. Life-style and mortality: a large-scale census-based cohort study in Japan. (Contributions to Epidemiology and Biostatistics, Vol 6). Basel: Karger, 1990
9. Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Hennekens CH, Speizer FE. Dietary fat and the risk of breast cancer. N Engl J Med 1987; 316: 22-8
10. Stellman SD, Garfinkel L. Smoking habits and tar levels in a new American Cancer Society Prospective study of 1.2 million men and women. JNCI 1986; 76: 1057-63
11. Brandt van den PA, Goldbohm RA, Veer van 't P et al., A large-scale prospective cohort study on diet and cancer in the Netherlands. J Clin Epidemiol 1990; 43: 285-95
12. Willett WC. Future directions in the development of food-frequency questionnaires. Am J Clin Nutr 1994; 59(suppl): 171-4s
13. Riboli E, Rönnholm H, Saracci R. Biological markers of diet. Cancer Surveys 1987; 6: 685-718
14. Beaton GH. Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology. Am J Clin Nutr 1994; 59(suppl): 253s-61s
15. Hunter D. Biochemical indicators of dietary intake. In: Willett W. Nutritional Epidemiology. Oxford: Oxford University Press, 1990, pp 143-216
16. Riboli E. Nutrition and cancer: Background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol 1992; 3: 783-91
17. Kaaks R, Plummer M, Riboli E, Estève J, Van Staveren WA. Adjustment for bias due to errors in exposure assessment in multi-center cohort studies on diet and cancer: a calibration approach. Am J Clin Nutr (suppl) 1994; 59: 245s-50s
18. Plummer M, Clayton D, Kaaks R. Calibration in multi-center cohort studies. Int J Epidemiol 1994; 23: 419-26
19. Marr JW. Individual dietary surveys: purposes and methods. World Rev Nutr Dietetics 1971; 13: 105-64
20. Block G. A review of validations of dietary assessment methods. Am J Epidemiol 1982; 115: 492-505
21. Bingham SA, Nelson M, Paul AA, Haraldsdottir J, Løken EB, Van Staveren WA. Methods for data collection at an individual level. In: Cameron ME, Van Staveren WA. Manual on methodology for food consumption studies. Oxford: Oxford University Press, 1988, pp. 53-106
22. Kaaks R, Riboli E, Estève J, Van Kappel AL, Van Staveren WA. Estimating the accuracy of dietary questionnaire assessment: validation in terms of structural equation models. Stat Med 1994; 13: 127-42
23. Burema J, Van Staveren WA, Van den Brandt PA. Validity and reproducibility. In: Cameron ME, Van Staveren WA. Manual on methodology for food consumption studies. Oxford: Oxford University Press, 1988, pp. 171-81
24. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press, 1992

25. Heitmann BL, Lissner L. Dietary underreporting by obese individuals - is it specific or non-specific? BMJ 1995; 311: 986-9
26. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of the exposure always bias a true effect toward the null value? Am J Epidemiol 1990; 132: 746-8
27. Walker AM, Blettner M. Comparing imperfect measures of exposure. Am J Epidemiol 1985; 121: 783-90
28. Marr JW, Heady JA. Within- and between-person variation in dietary surveys: number of days needed to classify individuals. Hum Nutr: Appl Nutr 1986; 40A: 347-64
29. Klerk de NH, English DR, Armstrong BK. A review of the effects of random measurement error on relative risk estimates in epidemiological studies. Int J Epidemiol 1989; 18: 705-12
30. Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement error on planning sample size required in a cohort study. Am J Epidemiol 1990; 132: 1185-95
31. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates on confidence intervals for systematic within-person measurement error. Stat Med 1989; 8: 1051-69
32. Bloemberg BPM. On the effect of measurement error in nutritional epidemiology using dietary history and food frequency methodology. Thesis. Leiden: State University, 1993
33. Willett W. Nutritional Epidemiology. Oxford: Oxford University Press, 1990
34. Hankin JH, Wilkens LR, Kolonel LN, Yoshizawa CN. Validation of a quantitative diet history method in Hawaii. Am J Epidemiol 1991; 133: 616-28
35. Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. Am J Epidemiol 1988; 128: 655-66
36. Nes M, Frost Andersen L, Solvoll K, Sandstad B, Hustvedt BE, Lovo A, Drevon CA. Accuracy of a quantitative food frequency questionnaire applied in elderly Norwegian women. Eur J Clin Nutr 1992; 46: 809-21
37. Nelson M. Assessment of food consumption and nutrient intake. Past intake. In: Margetts BM, Nelson M. Design concepts in nutritional epidemiology. Oxford: Oxford Medical Publications, 1991
38. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Garder L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986; 124: 453-69
39. Heady JA. Diets of bank clerks: development of a method of classifying the diets of individuals for use in epidemiologic studies. J R Stat Soc 1961; 124 (A): 336-61
40. Block G, Sinha R, Gridley G. Collection of dietary-supplement data and implications for analysis. Am J Clin Nutr 1994; 59(suppl): 232s-9s
41. Boutron MC, Faivre J, Milan C, Lorcerie B, Esteve J. A comparison of two diet history questionnaires that measure usual food intake. Nutr Cancer 1989; 12: 83-91
42. Tjonneland A, Haraldsdottir J, Overvad K, Stripp C, Ewertz M, Jensen OM. Influence of individually estimated portion size data on the validity of a semiquantitative food frequency questionnaire. Int J Epidemiol 1992; 21: 770-7
43. Haraldsdottir J. Minimizing error in the field: quality control in dietary surveys. Eur J Clin Nutr 1993; 47 (suppl 2): s19-s24
44. Willett WC, Sampson L, Browne ML, Stampfer MJ, Rosner B, Hennekens CH, Speizer FE. The use of a self-administered questionnaire to assess diet four years in the past. Am J Epidemiol 1988; 127: 188-99

45. West CE, Van Staveren WA. Food consumption, nutrient intake, and the use of food composition tables. In: Margetts BM, Nelson M. Design concepts in nutritional epidemiology. Oxford: Oxford Medical Publications, 1991
46. Mangels AR, Holden JM, Beecher GR, Forman MR, Lanza E. Carotenoid content of fruits and vegetables: an evaluation of analytic data. J Am Diet Assoc 1993; 93: 284-96
47. Pee de S, West CE. Lack of improvement in vitamin A status with increased consumption of dark-green vegetables. Lancet 1995; 346: 75-81
48. Vliet T van, Van den Berg H. ß-carotene absorption and cleavage in animals and man. A review. Eur J Clin Nutr, in press
49. Nelson M. The validation of dietary questionnaires. In: Nelson M. Design concepts in nutritional epidemiology. Oxford: Oxford Medical Publications, 1991
50. Beaton GH. Interpretation of results from diet history studies. In: Kohlmeier L, ed. The diet history method. Proceedings of the Second International Meeting on Nutritional Epidemiology. London: Smith-Gordon, 1989: 15-36
51. Flegal KM, Larkin FA. Partitioning macronutrient intake estimates from a food frequency questionnaire. Am J Epidemiol 1990; 131: 1046-58
52. Cameron ME, Van Staveren WA. Manual on methodology for food consumption studies. Oxford: Oxford University Press, 1988
53. Margetts BM, Nelson M. Design concepts in nutritional epidemiology. Oxford: Oxford Medical Publications, 1991
54. Bates CJ, Thurnham DI. Biochemical markers of nutrient intake. In: Margetts BM, Nelson M. Design concepts in nutritional epidemiology. Oxford: Oxford Medical Publications, 1991, 192-265
55. Comstock GW, Alberg AJ, Helzlsouer KJ. Reported effects of long-term freezer storage on concentrations of retinol, ß-carotene, and α-tocopherol in serum or plasma summarized. Clin Chem 1993; 39: 1075-8
56. Campbell DR, Gross MD, Martini MC, Grandits GA, Slavin JL, Potter JD. Plasma carotenoids as biomarkers of vegetable and fruit intake. Cancer Epidemiol Biom Prev 1994; 3: 493-500
57. Kardinaal AFM, Veer van 't P, Brants HAM, Berg van den H, Schoonhoven van J, Hermus RJJ. Relations between antioxidant vitamins in adipose tissue, plasma and diet. Am J Epidemiol 1995; 141: 440-50
58. Willett WC, Stampfer MJ, Underwood BA, Speizer FE, Rosner B, Hennekens CH. Validation of a dietary questionnaire with plasma carotenoid and alpha-tocopherol levels. Am J Clin Nutr 1983; 38: 631-9
59. Schäfer L, Overvad K. Subcutaneous adipose-tissue fatty acids and vitamin E in humans: relation to diet and sampling site. Am J Clin Nutr 1990; 52: 486-90
60. Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. Am J Clin Nutr 1985; 42: 1276-89
61. Isaksson B. Urinary nitrogen output as a validity test in dietary surveys. Am J Clin Nutr 1980; 33: 4-6
62. Beaton GH. What do we think we are measuring. In: Symposium on dietary data collection analysis and significance. Research Bulletin No. 675, Boston University of Massachusetts, 1982
63. Kaaks R, Riboli E, Van Staveren WA. Sample size requirements for calibration studies of dietary intake measurements in prospective cohort investigations. Am J Epidemiol 1995; 142: 557-65

64. BGA Commission on Nutritional Epidemiology. Recommendations for the design and analysis of nutritional epidemiologic studies with measurement errors in the exposure variables. Eur J Clin Nutr 1993; 47 (suppl): s53-7
65. Prentice AM (ed). The doubly labelled water method for measuring energy expenditure. International Atomic Energy Authority, Vienna, 1990
66. Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chron Dis 1978; 31: 399-418
67. Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models. Part I. Stat Med 1993; 12: 925-35

Summary

Although diet is known to play an important role in the development of cancer, the state of knowledge on this topic is still limited. Vegetables and fruits belong to the few dietary components for which the relationships with cancer are well established by epidemiologic studies, i.e. inverse associations are consistently observed for cancer at many sites. The strength of the associations between intake of vegetables and fruits and the risk of cancer at different sites is, however, still unknown. One major complicating factor to ascertain the strength is measurement error in data on vegetable and fruit intake. The antioxidant (pro)vitamins ß-carotene, vitamin C, and vitamin E, are three of many substances in vegetables and fruits which may be responsible for the anticarcinogenic effect. The problem of assessing intake of vegetables, fruits, and antioxidant (pro)vitamins is therefore the central focus of this thesis.

In the first part of the thesis, two studies on vegetables, fruits, antioxidant (pro)vitamins and cancer are described. A lack of information on the extent of measurement error in the dietary intake data in both studies hampered the correct interpretation of the results.

In chapter 2, differences in the average intakes of antioxidant (pro)vitamins were studied in relation to differences in population mortality rates from lung, stomach and colorectal cancer among the 16 cohorts of the Seven Countries Study. In the 1960s detailed dietary information was collected in small subsamples of the cohorts with the dietary record method. Food equivalent composites representing the average food intake of each cohort at baseline were later collected locally and analyzed in a central laboratory. The average intake of vitamin C was strongly inversely related to 25-year stomach cancer mortality rates ($r=-0.66$, $p=0.01$). Adjustment for smoking and intake of salt or nitrate did not alter the results. The average intakes of α-carotene, ß-carotene, and α-tocopherol were not independently related to mortality rates from lung, stomach, or colorectal cancer, nor was vitamin C related to lung and colorectal cancer. The possibility of biased measures of association because of other factors that may have confounded the associations or because of dietary data which may have been differentially biased across the cohorts could not be ruled out.

The intake of vegetables, fruits, ß-carotene, vitamin C, and vitamin E in relation to incidence of lung cancer at the individual level was described in chapter 3. For 561 men from the town of Zutphen, the Netherlands, dietary history information was obtained in 1960, 1965, and 1970. During 1971-1990 54 new cases of lung cancer

were identified. No relationship between the intake of vitamin E and lung cancer risk was observed. For vitamin C intake the results pointed to an inverse association, although not entirely consistently. Furthermore, it was observed that subjects with low stable intakes (i.e. low in 1960, 1965, and 1970) of vegetables, fruits, and ß-carotene experienced more than two-fold increased relative risks on lung cancer than those with high stable intakes. Relative risks for subjects with low versus high average intake were much lower. Considerable measurement error in dietary assessment could be expected in this study since intake of vegetables and fruits was only assessed crudely in the 1960s. Together with the possibility of residual confounding by smoking, it was not possible to draw definite conclusions from this study.

The second part of the thesis includes several studies on the estimation of measurement error in data on vegetable, fruit, and antioxidant (pro)vitamin intake and biochemical markers.

In chapter 4, the effects of frozen storage at -20 °C on (pro)vitamin concentrations in EDTA-plasma and whole blood was studied. Aliquots from 55 samples were analyzed before storage and after 3, 6, 12, 24, 36 and 48 months at -20 °C. Dramatic decreases occurred for EDTA-plasma concentrations of vitamin E between 6 and 12 months, vitamin A, total carotenoids and ß-carotene after 1 year, and whole blood niacin after 4 years. A smaller decrease was observed for folic acid at 1 year of storage, but the level remained constant thereafter. The vitamins D, B_6, B_{12} (EDTA-plasma), B_1 and B_2 (whole blood) showed no decline during 4 years of storage. With the exception of folic acid, the observed decreases varied considerably among subjects. In a simulation study it was shown that using vitamin concentrations in blood stored at -20 °C results in highly attenuated odds ratios for the instable vitamins like ß-carotene and vitamin E. For prospective studies on these (pro)vitamins storage at -80 °C or colder is required and a stability study should form part of the study.

A self-administered food frequency questionnaire developed for the Dutch cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC) is described in chapter 5. The questionnaire contains photographs to estimate portion sizes of 28 food items, and habitual consumption of 178 food items can be calculated from the questionnaire data. Reproducibility and relative validity for food group and nutrient intake were investigated in a population of 121 Dutch men and women, as reported in chapters 5 and 6. To assess the relative validity 12 monthly 24-h recalls served as

reference method together with several biochemical markers of intake. For vegetables, fruits, and antioxidant (pro)vitamins, reproducibility after 6 months ranged from 0.70 to 0.81 in men and from 0.61 to 0.77 in women. Spearman correlation coefficients between estimates of vegetable intake based on the questionnaire and those based on 24-h recalls were 0.38 among men and 0.31 among women. For fruit intake the correlation coefficients were 0.68 and 0.56, whereas the median correlation coefficients for all food groups were 0.61 and 0.53, respectively. The median of Pearson correlation coefficients between nutrient intakes assessed by the questionnaire and 24-h recalls was 0.59 for men with correlations of 0.26 for ß-carotene, 0.39 for vitamin C, and 0.57 for vitamin E. For women the median was 0.58 with values of 0.35, 0.58, and 0.44 for the three antioxidant (pro)vitamins respectively. Associations with serum ß-carotene (r=-0.16 for men; 0.13 for women) and α-tocopherol (0.23 and 0.15, respectively) were worse than those obtained with 24-h recalls. The questionnaire seemed adequate for ranking Dutch EPIC subjects according to intake of most food groups including fruits, although the relative validity for some food groups, such as vegetables, was poor. Concerning nutrients, the questionnaire seemed adequate for ranking subjects according to intake of energy, macronutrients, dietary fibre and retinol, but it did not yield such good results for ß-carotene, vitamin C for men, and vitamin E for women. The observed correlation coefficients may be either over- or underestimates of the true validity coefficients, because of unknown error structure in the questionnaire and reference methods.

An alternative way to estimate validity coefficients, by a triangular comparison between questionnaire, reference and biomarker measurements is presented in chapter 7. This so-called method of triads assumes that the measurements are linearly related to true intake and have independent random errors. The method was applied to examples from the EPIC-study. In some examples 'Heywood cases' occurred, i.e. estimated validity coefficients greater than one, or validity coefficients which could not be estimated. This can be caused by random sampling fluctuations or by violation of the model assumptions. One likely violation is a positive correlation between the random errors of questionnaire and reference measurements which would result in overestimated validity coefficients. Confidence intervals of the validity coefficients were estimated by using a bootstrap method. Validity studies with several hundreds of subjects and/or more accurate biochemical indicators of dietary intake are needed to

estimate validity coefficients precisely and avoid complications with the bootstrap method.

From these studies it is concluded in chapter 8, that measurement error in assessing vegetable, fruit, and antioxidant (pro)vitamin intake may be large which is a handicap for epidemiologic studies. Random measurement error frustrates the power for hypothesis testing, and both random and systematic measurement error affect the assessment of the strength of an association. Further progress lies in improvement and adaptation of dietary assessment methods to specific aims, but probably even more in understanding error structures and analytical methods to recognize and cope with that structure.

Samenvatting

Hoewel het bekend is dat voedingsfactoren een belangrijke rol spelen bij de ontwikkeling van kanker, is er nog veel onduidelijk op dit gebied. Groenten en fruit behoren tot de weinige componenten in de voeding waarvoor consistente resultaten gevonden worden in epidemiologisch onderzoek. Voor deze produktgroepen worden inverse relaties gerapporteerd met verschillende vormen van kanker; de sterkte van de verbanden is echter onduidelijk. Eén factor waardoor onderzoek op dit gebied bemoeilijkt wordt, zijn meetfouten in de schatting van de groente- en fruitconsumptie. De antioxidant (pro)vitamines ß-caroteen, vitamine C en vitamine E, zijn drie van de vele stoffen in groenten en fruit die mogelijk verantwoordelijk zijn voor het anticarcinogene effect. Het nagaan van de inneming van groenten, fruit en antioxidant (pro)vitamines vormt daarom het centrale thema van dit proefschrift.

In het eerste deel van het proefschrift worden twee onderzoeken beschreven naar de relatie tussen groenten, fruit, antioxidant (pro)vitamines en kanker. Een juiste interpretatie van de resultaten van deze onderzoeken is echter moeilijk doordat geen informatie aanwezig is over de meetfout in de voedingsgegevens.

In hoofdstuk 2 werden verschillen in de gemiddelde inneming van antioxidant (pro)vitamines bestudeerd in relatie tot verschillen in sterftecijfers aan long-, maag-, en colorectaal-kanker in de 16 cohorten van de Zeven Landen Studie. In kleine steekproeven van deze cohorten is in de zestiger jaren de voedselconsumptie nagegaan door middel van de opschrijfmethode. Later zijn voedingsmiddelen, die de gemiddelde inneming van ieder cohort weergeven, lokaal gekocht en vervolgens in een laboratorium chemisch geanalyseerd. De gemiddelde inneming van vitamine C was sterk invers gerelateerd aan de maagkankersterfte in de 25 jaar erna ($r=-0.66$; $p=0.01$), ook na correctie voor roken en de zout- of nitraatinneming. De gemiddelde innemingen van α-caroteen, ß-caroteen, en α-tocoferol waren niet onafhankelijk gerelateerd aan long-, maag-, of colorectaal-kanker. Ook was de gemiddelde inneming van vitamine C niet gerelateerd aan long- of colorectaal-kanker. Het kan echter niet worden uitgesloten dat de gevonden resultaten vertekend zijn als gevolg van andere factoren waarvoor niet gecorrigeerd is of door cohort-specifieke systematische meetfouten in de voedingsgegevens.

Een onderzoek naar de relatie tussen de inneming van groenten, fruit, ß-caroteen, vitamine C en vitamine E en het optreden van longkanker op individueel niveau is beschreven in hoofdstuk 3. In 1960, 1965, en 1970 zijn bij 561 mannen uit

Zutphen voedingsgegevens verzameld met de 'dietary history' methode. In de periode 1971-1990 werden 54 nieuwe gevallen van longkanker waargenomen. Er werd geen relatie gevonden tussen de inneming van vitamine E en het risico op longkanker. Voor vitamine C wezen de resultaten op een inverse relatie, hoewel dit verband niet geheel consistent was. Verder bleken mannen met een stabiel lage inneming (laag in 1960, 1965 en 1970) van groenten, fruit, en ß-caroteen een meer dan tweevoudig hoger risico te hebben op het ontwikkelen van longkanker dan mannen met een stabiel hoge inneming. De relatieve risico's voor mannen met een gemiddeld lage versus gemiddeld hoge inneming waren veel zwakker. Een beperking van dit onderzoek is dat de gegevens van de groente- en fruitinneming waarschijnlijk niet zo valide zijn, door de grove wijze van navraag in de zestiger jaren. Samen met de mogelijkheid dat er niet volledig gecorrigeerd kon worden voor het effect van roken, bemoeilijkt dit het trekken van juiste conclusies.

In het tweede deel van het proefschrift worden enkele onderzoeken beschreven naar meetfouten in gegevens over de inneming van groenten, fruit en antioxidant (pro)vitamines en hun biochemische merkers.

De effecten van opslag bij -20 °C op concentraties van (pro)vitamines in EDTA-plasma en volbloed zijn beschreven in hoofdstuk 4. Kleine hoeveelheden van 55 bloedmonsters werden geanalyseerd voordat ze ingevroren werden en ook na 3, 6, 12, 24, 36, en 48 maanden bij -20 °C. Een groot verval trad op tussen 6 en 12 maanden voor vitamine E, na 1 jaar voor vitamine A, totaal carotenoïden, en ß-caroteen (allen in EDTA-plasma) en na 4 jaar voor nicotinezuur (in volbloed). Een geringer verval werd na 1 jaar waargenomen voor foliumzuur, maar de concentraties bleven daarna gelijk. De vitamines D, B_6, B_{12} (EDTA-plasma), B_1 en B_2 (volbloed) waren gedurende 4 jaar bij -20 °C stabiel. Met uitzondering van foliumzuur, werden grote tussenpersoonsverschillen waargenomen in de mate van verval van de instabiele (pro)vitamines. Met behulp van een simulatiestudie werd geïllustreerd dat het gebruik van concentraties van ß-caroteen en vitamine E in EDTA-plasma dat bij -20 °C bewaard is, resulteert in sterk verzwakte odds ratio's. Het wordt daarom voor prospectieve studies aanbevolen om bloed bij -80 °C of kouder op te slaan en om altijd een stabiliteitsonderzoek uit te voeren.

In hoofdstuk 5 wordt een schriftelijke voedselfrequentievragenlijst beschreven die ontwikkeld is voor het Nederlandse cohort van de Europese Prospectieve Studie

naar Voeding en Kanker (EPIC). De vragenlijst bevat foto's voor het schatten van portiegroottes van 28 voedingsmiddelen. Door middel van deze vragenlijst kan de gewoonlijke consumptie van 178 voedingsmiddelen berekend worden. In de hoofdstukken 5 en 6 zijn de resultaten beschreven van een onderzoek naar de herhaalbaarheid en relatieve validiteit van deze vragenlijst bij een populatie van 121 Nederlandse mannen en vrouwen. Twaalf maandelijkse 24-uurs 'recalls' en een aantal biomerkers dienden als referentiemethoden om de relatieve validiteit na te gaan. De 6-maands herhaalbaarheid voor groenten, fruit, en antioxidant (pro)vitamines varieerde bij mannen van 0.70 tot 0.81 en bij vrouwen van 0.61 tot 0.77. Spearman correlatiecoëfficiënten tussen schattingen van de groente-inneming op basis van de vragenlijst en herhaalde 24-uurs 'recalls' waren 0.38 voor mannen en 0.31 voor vrouwen. Voor de fruitinneming waren deze correlaties respectievelijk 0.68 en 0.56 terwijl de medianen van de coëfficiënten voor alle voedingsmiddelengroepen respectievelijk 0.61 en 0.53 waren. Voor voedingsstoffen was de mediaan van de Pearson correlatiecoëfficiënten 0.59 voor mannen, met correlaties van 0.26 voor ß-caroteen, 0.39 voor vitamine C en 0.57 voor vitamine E. Voor vrouwen was de mediaan 0.58 met waarden van respectievelijk 0.35, 0.58 en 0.44 voor de drie antioxidant (pro)vitamines. Correlaties met serumconcentraties van ß-caroteen (r=-0.16 voor mannen; 0.13 voor vrouwen) en α-tocoferol (0.23 voor mannen en 0.15 voor vrouwen) waren veel lager dan die met de herhaalde 24-uurs 'recalls'. De voedselfrequentievragenlijst leek voldoende in staat om de Nederlandse EPIC-deelnemers te ordenen van een kleine naar een grote inneming van de meeste voedingsmiddelengroepen. Dit gold ook voor fruit maar niet voor groenten. Het ordenen van mensen op basis van hun voedingsstoffeninneming ging redelijk goed voor energie, macronutriënten, voedingsvezel en retinol, terwijl minder goede resultaten verkregen werden voor ß-caroteen, vitamine C bij mannen en vitamine E bij vrouwen. De gevonden correlatiecoëfficiënten kunnen echter over- of onderschattingen zijn van de werkelijke validiteitscoëfficiënten, omdat de foutenstructuur in vragenlijst- en referentiegegevens onbekend is.

In hoofdstuk 7 werd een andere manier beschreven voor het schatten van validiteitscoëfficiënten. Deze maakt gebruik van de driehoeksvergelijking tussen de vragenlijst-, 24-uurs 'recall'- en biomerkergegevens en wordt de triademethode genoemd. De aanname bij deze methode is dat alle metingen lineair met de werkelijk inneming samenhangen, en dat de drie methoden onafhankelijke meetfouten hebben.

De triademethode werd toegepast op voorbeelden van de EPIC-studie. In enkele voorbeelden traden 'Heywood'-gevallen op, waarbij de validiteitscoëfficiënten niet geschat konden worden of een schatting van groter dan 1 opleverden. Dit kan veroorzaakt zijn door willekeurige steekproeffluctuaties of doordat de modelaannames niet correct waren. Een positieve correlatie tussen de willekeurige meetfouten in de vragenlijst en de 24-uurs 'recalls' is een mogelijke schending van de modelaannames, die een overschatting van de validiteitscoëfficiënt zou opleveren. Betrouwbaarheidsintervallen van de validiteitscoëfficiënten werden geschat met de 'bootstrap'-methode. Om nauwe betrouwbaarheidsintervallen te krijgen en problemen met de 'bootstrap'-methode te voorkomen zijn validiteitsonderzoeken met enkele honderden deelnemers nodig en/of nauwkeurige biomerkers van de voedingsinneming.

In hoofdstuk 8 wordt uit voorgaande onderzoeken geconcludeerd dat meetfouten in consumptiegegevens over groenten, fruit en antioxidant (pro)vitamines groot kunnen zijn waardoor epidemiologisch onderzoek op dit terrein complex is. Door willekeurige meetfouten is het moeilijk na te gaan of een associatie al dan niet aanwezig is, terwijl het bepalen van de sterkte van een relatie bemoeilijkt wordt door zowel willekeurige als systematische meetfouten. Vooruitgang op dit gebied moet gezocht worden in het verder ontwikkelen en aanpassen van voedselconsumptie-methoden, maar waarschijnlijk nog meer in het begrijpen van foutenstructuren in voedingsgegevens en het ontwikkelen van analysemethoden om deze structuren te herkennen en te hanteren.

Dankwoord

Dit proefschrift is het resultaat van een leuke werkperiode. Inhoudelijk leuk door het onderwerp, maar net zo belangrijk, zeer plezierig door de samenwerking met begeleiders en collega's. Het aantal mede-auteurs van de verschillende hoofdstukken, ik tel er maar liefst 22, geeft aan dat veel mensen aan dit proefschrift hebben meegewerkt. Achter de schermen hebben nog tientallen mensen aan de diverse studies bijgedragen, zoals aan de gegevensverzameling, de administratieve, PC- en software-ondersteuning en het scheppen van een plezierig werkklimaat. Ik kan ze hier niet allemaal noemen, maar ben hun allen zeer erkentelijk. Verder ben ik dank verschuldigd aan de deelnemers van de verschillende studies; zonder hen is dit type onderzoek niet mogelijk. Een aantal personen wil ik speciaal noemen.

Allereerst wil ik mijn (co)promotoren, Bas Bueno de Mesquita, Daan Kromhout en Wija van Staveren, bedanken. Met name in het laatste stadium heb ik in korte tijd veel van jullie gevraagd, in verband met de 'andere' bevalling. Bas, in 1990 ben ik aan een lege tafel bij jou op de kamer aan dit karwei begonnen: het was toen nog niet duidelijk dat het een proefschrift zou worden. Ook nu zijn we nog kamergenoten en dat ervaar ik als zeer plezierig. Ik heb veel geleerd van je grondige en integere werkwijze. Daan, jij bood me de mogelijkheid om een proefschrift te schrijven en stelde de gegevens van de stabiliteitsstudie, de Zeven Landen Studie en de Zutphen Studie hiervoor beschikbaar. Ik ben niet de eerste die het in zijn/haar dankwoord schrijft, maar het is ongelofelijk hoe je binnen je drukke funktie kans ziet om concept-artikelen uiterst snel en grondig door te nemen en van kommentaar te voorzien. Wija, hartelijk dank voor je enthousiaste begeleiding. Jouw werk op het gebied van voedselconsumptiemethoden is mij altijd tot voorbeeld geweest.

Van de diverse studies die in dit proefschrift beschreven staan, ben ik het meest betrokken geweest bij het BALANS-onderzoek, omdat ik dit vanaf de subsidie-aanvraag tot de rapportage (hoofdstuk 5 en 6) heb meegemaakt. Ik wil Ester Goddijn, Annemarie Jansen en Margreet Pols hartelijk danken voor de leuke samenwerking en hun grote inzet om de dagelijkse gang van zaken van de gegevensverzameling in goede banen te leiden. Bas Bueno de Mesquita, Gatske Obermann-de Boer, Petra Peeters, en Jet Smit hebben er zorg voor gedragen dat de opzet en gegevensverzameling van de studie een goed evenwicht behield tussen wat wetenschappelijk wenselijk en praktisch haalbaar was.

Na het BALANS-onderzoek, een pilot-studie van het EPIC-onderzoek, ben ik betrokken gebleven bij het voedingsgedeelte van het EPIC-hoofdonderzoek. Ik heb het altijd erg prettig gevonden om het achter de PC zitten in verband met mijn proefschrift

af te wisselen met deze taken. Dit komt onder meer door de gezellige en constructieve samenwerking met de andere leden van de voedingswerkgroep, nu bestaande uit Sigrid Braak, Bas Bueno de Mesquita, Ester Goddijn en Janneke Verloop. Het feit dat de EPIC-studie een internationale studie is, bood me de gelegenheid om samen te werken met Europese collega's, met name in de werkgroep voor de standaardisatie van het 24-uurs recall interview in het kader van de calibratiestudie. Dit was een zeer nuttige en leuke ervaring. I would like to thank Elio Riboli, the principal investigator of the EPIC-study, who gave me the opportunity to work for a short time at his unit of the International Agency for Research on Cancer in Lyon. Hier heb ik samen met Rudolf Kaaks gewerkt aan hoofdstuk 7 van dit proefschrift. Rudolf, dank voor de vele leerzame discussies, vaak tot laat in de avond, en de gastvrijheid van jou en Franca. Jouw proefschrift heeft volgens mij een grote bijdrage geleverd aan de voedingsmethodologie en ik heb geprobeerd delen ervan in praktijk te brengen.

De datasets van de stabiliteitsstudie, de Zeven Landen Studie en de Zutphen Studie waren kant- en klaar voor mij beschikbaar om statistische analyses uit te voeren en hierover artikelen te schrijven. Ik weet dat er heel veel bij komt kijken voor het zover is, en wil iedereen die hieraan bijgedragen heeft hartelijk danken. De directie van het RIVM ben ik erkentelijk voor de geboden gelegenheid om de resultaten van de studies te bewerken tot een proefschrift binnen de onderzoekschool NIHES.

Mijn maatje Tom was onmisbaar voor het delen van en meedenken met alles, zowel de ups als de downs. Ook dank aan mijn ouders, overige familie en vrienden die zorgden voor de broodnodige afleiding. Mijn paranimfen Margreet Pols en Trinette van Vliet wil ik danken voor het meeleven tot nu toe en hun bijdrage die, op het moment dat ik dit schrijf, nog moet komen.

Curriculum vitae

Marga Ocké werd geboren op 7 november 1964 te IJzendijke, gemeente Oostburg. In 1983 behaalde zij het Atheneum diploma aan de scholengemeenschap St. Eloy te Oostburg. Datzelfde jaar begon zij aan de studie 'Voeding van de Mens' aan de Landbouwuniversiteit Wageningen. Zij behaalde in 1989 het ingenieursdiploma met afstudeervakken Voedingsleer, Gezondheidsleer en Statistiek. Na haar studie heeft zij tijdelijk werk verricht bij het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) te Bilthoven en op het National Institute of Nutrition te Hanoi, Vietnam. Vanaf 1 juni 1990 is zij weer bij het Centrum voor Chronische Ziekte en Milieu Epidemiologie van het RIVM werkzaam. Na haar betrokkenheid bij de voorbereiding en uitvoering van een validatiestudie, werd haar de mogelijkheid geboden om een proefschrift te schrijven met de gegevens van deze studie en andere beschikbare datasets.