

Complex correlation structure in consumption rates of major food groups: implications for the assessment of radiation exposure

Melanie I Breuninger¹, Gabriel A Schachtel², Petra M Lührmann¹, Bernd Hartmann¹ and Monika Neuhäuser-Berthold^{1,*}

¹Institute of Nutritional Science, Justus-Liebig-University of Giessen, Goethestrasse 55, D-35390 Giessen, Germany;

²BioStatistics, Justus-Liebig-University of Giessen, Germany

Submitted 13 November 2001; Accepted 13 June 2002

Abstract

Objective: An extensive dataset on individual food consumption was analysed in order to study all pairwise correlations between the consumption rates of 11 major food groups. Additionally, the project aimed to examine and quantify the accuracy of a recently proposed estimator of total food consumption to be used for the estimation of radiation exposure by food. Such an inquiry seems justified, because the proposed estimator implicitly presumes an essentially positive correlation structure in food consumption rates.

Design: Statistical analysis using representative data gathered in Germany in a nationwide food consumption survey.

Setting: Germany.

Subjects: Individuals aged between 4 and 94 years namely 10 901 males and 12 308 females.

Results: The consumption rates of 11 major food categories showed several significantly positive, but also a number of significantly negative, correlations. Negative associations between cereal and potato products persisted consistently over all age groups, independent of sex. Other significantly negative correlations were limited to certain age groups. Reflecting these negative correlations, a subsequent analysis of relative ranks of consumption revealed that no person in the sample had the highest consumption rates in all food groups simultaneously. Based on representative samples, overestimations of 34 to 53% were obtained if – as recently suggested in the context of radiation exposure prediction – the 95th percentiles of total food consumption were determined as sums of the corresponding percentiles of the food groups.

Conclusions: The complex correlation structure of food group consumption rates, as identified in this study, bears important implications for various health-related issues. Ignoring them could lead to overly conservative estimations of radiation exposure due to food ingestion or to confounding effects in epidemiological studies on nutritional risk factors of diseases. The results also indicate that a distinction into different dietary patterns might be useful in characterising different consumption habits.

Keywords

Correlation between intake of food groups
Radiation exposure prediction
Unbiased estimation

It is a well-known fact that different components of nutritional intake can substitute for each other, while others tend to complement one another. A number of publications describe specific relationships between certain food groups. For example, Ursin *et al.*¹ reported about circumstances where fruits and vegetables substituted for fat consumption and Elmstahl *et al.*² observed associations between meat consumption and various other food groups. However, none of these studies provides a systematic analysis of the entire correlation structure of food group consumption rates.

One aim of the present paper is therefore to attempt a

thorough investigation of all pairwise correlations between 11 major food groups. It is based on a large dataset gathered for Nationale Verzehrsstudie (NVS), the first nation-wide food consumption survey in Germany³. This survey was commissioned and supported by the Ministry of Science and Technology and is representative of the West German population living in private households (before the 1990 reunification of Germany). The data were collected between 1985 and 1989, involving a total of 24 632 individuals. Among other things, participants carried out a 7-day food record and reported on sociodemographic data.

*Corresponding author. Email monika.neuhaeuser-berthold@ernaehrung.uni-giessen.de

Our interest in the interdependencies between consumption rates of food groups was originally prompted by questions related to the assessment of radiation exposure of man from releases of radionuclides into the environment. The Council Directive 96/29/EURATOM⁴ is laying down basic safety standards for the radiation protection of the general public. In Article 45 it demands of competent authorities to ensure that dose estimates are being made as realistically as possible. For Germany, the recent Radiation Protection Ordinance (RPO)⁵ specifies the measures and estimation methods to be taken to achieve these safety standards. A second objective of our paper is therefore to evaluate the validity of the proposed methods.

The Guideline to the German RPO⁶ as issued in 1990 applies the following formula for estimation of the yearly radiation dose $H_{T,r}$ ingested by food. With respect to a certain tissue T and nuclide r , the dose $H_{T,r}$ is defined there as

$$H_{T,r} = (U^{\text{Bl}} \times C_r^{\text{Bl}} + U^{\text{Pf}} \times C_r^{\text{Pf}} + U^{\text{Mi}} \times C_r^{\text{Mi}} + U^{\text{Fl}} \times C_r^{\text{Fl}}) \times g_{T,r}, \quad (1)$$

or

$$H_{T,r} = \left(\sum U^{(i)} \times C_r^{(i)} \right) \times g_{T,r},$$

where Bl, Pf, Mi and Fl denote the four food groups: 'leaf vegetable', 'other plant products', 'milk products' (including milk) and 'meat products' (including meat), respectively. Furthermore, for any food group (i), the coefficients $U^{(i)}$ and $C_r^{(i)}$ denote the consumption per year (kg) and the specific activity of nuclide r ($\text{Bq} \times \text{kg}^{-1}$), respectively. Finally, $g_{T,r}$ is the dose factor for ingestion of nuclide r ($\text{Sv} \times \text{Bq}^{-1}$) in tissue T .

Thus, in the original Guideline, the yearly radiation dose due to ingestion should be obtained by summing up the specific contributions of only four food groups to the radionuclide intake. For radioecological reasons, the recent RPO⁵ from 2001 expands the number of food groups distinguished from four to nine by splitting the group 'vegetables' into four new groups, namely 'cereals', 'fruits', 'roots' and 'other vegetables'. Additionally, 'fish' is considered for the aquatic pathway and 'human breast milk' for infants.

The RPO⁵ in Section 47(2) together with Appendix VII regulates which food consumption rates should be used to estimate the radiation ingested by food. It is recommended there to use the 95th percentiles of the consumption rates as $U^{(i)}$ values in equation (1). However, without knowledge about the underlying correlation structure of the food groups involved, the RPO approach seems rather arbitrary. Only under very restricted conditions one can expect this method to supply an unbiased estimation for the 95th percentile of total food consumption (or for the corresponding radioactivity), e.g. for normally distributed consumption rates within each food group only perfect (positive) pairwise correlations will yield an unbiased estimation (see Appendix). In general, though, especially

when some food groups are negatively correlated to each other, a pronounced overestimation is most likely to occur. We therefore employed the NVS consumption data in order to explore the existing correlation structure and to quantify empirically the actual extent of bias attached to the RPO estimation method.

Methods

Study design

The study was carried out by the Institute of Nutritional Science at the Justus-Liebig-University of Giessen on behalf of the German National Office for Radiation Protection. The analysis was based on a dataset that was retrieved from the Public Use File⁷ of the NVS³. As requested by the German National Office for Radiation Protection, food groups were specified following the food groups formed in the supplementary report of the NVS⁸. In deviation from the RPO Guideline we therefore differentiated the universe of nutritional intake into 11 rather than four or nine food groups, namely: (1) 'meat products' (including meat); (2) 'fish'; (3) 'eggs'; (4) 'milk products' (including milk); (5) 'pastries' (including bread and cakes); (6) 'cereal products'; (7) 'potato products' (including potatoes); (8) 'fruits'; (9) 'non-leaf vegetables'; (10) 'leaf vegetables'; and (11) 'vegetable products'.

Subjects

Data on food consumption and sociodemographic characteristics of 10 901 male and 12 308 female individuals aged between 4 and 94 years were used.

Statistical methods

Statistical analyses were carried out with SPSS for Windows, Version 9.0. Where normality could not be assumed, medians were used as location parameters, and the 5th and 95th percentiles as a measure of variation. Spearman correlation coefficients were calculated to determine the pairwise association between food groups. Coefficients were considered statistically significant if their P -value fell below 0.05.

To investigate the distribution of total food consumption, we calculated, for each participating individual $i = 1, 2, \dots, N$, the total relative rank TRR _{i} in the following way. First we assigned to each participant i in each food group $j = 1, 2, \dots, 11$ a relative rank $r_{ij} = R_{ij}/N$, where R_{ij} is the participant's rank with respect to his/her consumption rate in food group j . In each food group the person with the lowest (or highest) consumption rate received the smallest (or largest) possible relative rank, namely $1/N$ (or $N/N = 1$). By summation over all 11 relative ranks of participant i , one obtains the total relative rank

$$\text{TRR}_i = \sum_{j=1}^{11} r_{ij}$$

as a measure of that person's total consumption. Note that, because of $0 < r_{ij} \leq 1$, we get $0 < TRR_i \leq 11$. The theoretical TRR maximum of 11, however, is only attained if the same individual exhibits, for all 11 food groups simultaneously, the highest consumption rates in the corresponding sample.

To measure the extent of bias of the RPO proposed 95th percentile estimator we calculated a 'factor of over-estimation' as the ratio of the sample's estimated value over the sample's actual 95th percentile. To assess statistical significance we determined the 99.9% confidence interval for the sample's 95th percentile.

Results

The data on the daily consumption rates of the different food groups are summarised in Tables 1 and 2 for men and women, respectively. For each food category, the median and the 5th and 95th percentiles of the absolute amount (g) are listed for nine age groups. The age classification partitions childhood into the three groups, 4–6, 7–9 and 10–12 years, adolescence into 13–14 and 15–18 years, adulthood into 19–24, 25–50 and 51–64 years, and, finally, seniors in the class more than 64 years.

A comparison of the age-dependent development of medians exhibits only minor differences between the sexes but substantial non-parallelism between different food categories. With increasing age the medians in some food groups tend towards steady growth (e.g. vegetables, potatoes, pastries), others show nearly continuous decline (e.g. milk products) and others again (e.g. cereals) increase first towards a peak before declining after a certain age. Fruit consumption declines first to reach a minimum in early adulthood and then increases in late adulthood and seniority. These brief and crude comparisons already indicate the existence of complex correlation patterns with positive as well as negative and even non-significant associations.

The following correlation analysis strongly confirms these first impressions. Its most important results are summarised in Tables 3 and 4, where only statistically significant correlations are listed, with all negative ones highlighted by bold print. Table 3 contains all coefficients involving meat products, and Table 4 all those involving cereal products.

In both tables most entries show positive correlations. In one case, namely for meat products and pastries, the correlations are – without exception – positive for all age groups and both sexes. With only minor exceptions the same holds for the correlations between meat and potato products, and between meat products and all three vegetable groups (Table 3). Positive correlations exist independent of sex also between cereal products and milk products (for 13 years and older); between cereal products and eggs (for 15 years and older); and

between cereal products and fruits (for 19 years and older), see Table 4.

Negative correlations were less frequent, nevertheless one association was – independent of age and sex – consistently negative, namely the correlation between cereal products and potato products. Children showed negative correlations between the rates of cereals and pastries as well as between cereal products and meat products. For men and women above the age of 15 years, negative correlations were observed between meat products and milk products. Finally, associations between fish and cereal products, fish and meat products, as well as fruit and meat products, tended to be negative in adults.

In all age groups and in both sexes, the distributions of the total relative ranks of individual food ingestion were remarkably similar in central tendency, symmetry and unimodality, suggesting nearly normal distributions. The typical shape of the corresponding histograms is displayed

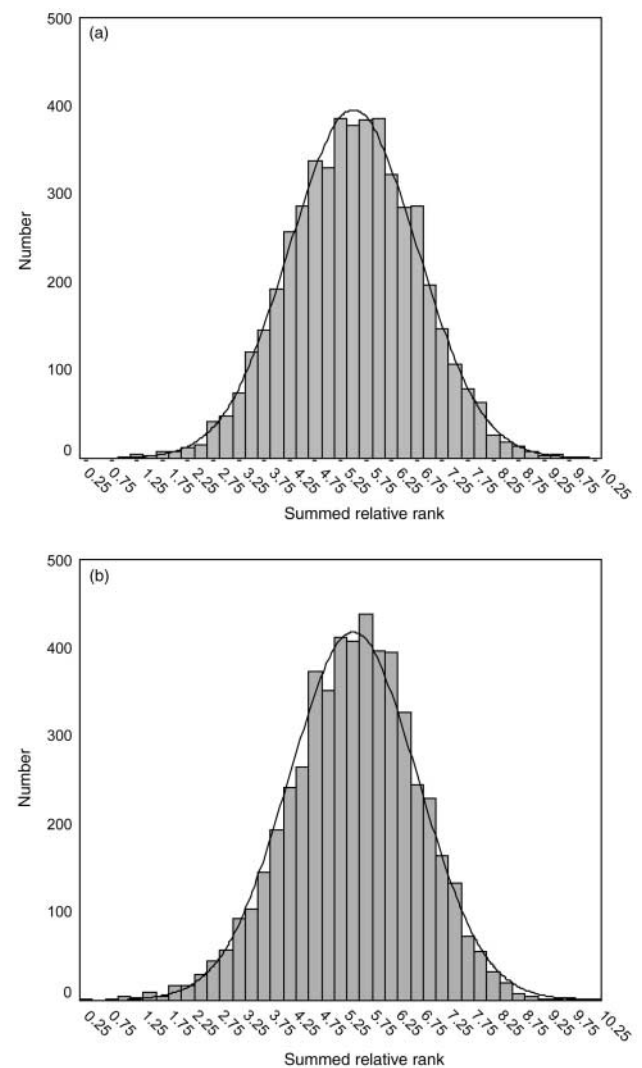


Fig. 1 Frequency distribution of the summed relative ranks of food ingestion of the 25–50-year-old subjects: (a) male subjects; (b) female subjects

Table 1 Daily consumption rates of the different food groups (g) in men (median, 5th–95th percentiles)

| | Age group (years) | | | | | | | | | |
|---------------------|-------------------|--------------|---------------|---------------|---------------|----------------|----------------|----------------|---------------|--|
| | 4–6 (n=450) | 7–9 (n=441) | 10–12 (n=358) | 13–14 (n=275) | 15–18 (n=592) | 19–24 (n=1182) | 25–50 (n=4974) | 51–64 (n=1699) | > 64 (n=930) | |
| Fish | 0 (0–29) | 0 (0–33) | 0 (0–39) | 0 (0–43) | 0 (0–54) | 0 (0–58) | 8 (0–70) | 14 (0–77) | 16 (0–83) | |
| Meat products | 80 (23–160) | 105 (35–208) | 131 (53–244) | 145 (61–279) | 172 (68–348) | 183 (48–363) | 188 (70–355) | 181 (82–322) | 153 (61–289) | |
| Eggs | 18 (2–49) | 22 (3–58) | 21 (2–54) | 24 (2–66) | 27 (3–70) | 29 (2–83) | 30 (3–84) | 29 (3–83) | 30 (2–75) | |
| Milk products | 302 (75–699) | 322 (78–764) | 346 (62–748) | 328 (64–834) | 287 (38–895) | 211 (31–784) | 138 (24–543) | 134 (22–511) | 147 (27–543) | |
| Pastries | 132 (63–237) | 164 (70–302) | 185 (76–415) | 216 (93–500) | 238 (89–462) | 225 (96–462) | 235 (96–459) | 243 (109–460) | 240 (106–438) | |
| Cereal products | 31 (7–92) | 41 (7–121) | 40 (6–127) | 44 (10–136) | 39 (5–137) | 39 (8–126) | 31 (5–105) | 26 (4–90) | 23 (4–75) | |
| Potato products | 63 (14–156) | 74 (14–183) | 91 (24–211) | 100 (21–253) | 107 (27–249) | 105 (10–262) | 103 (20–254) | 121 (27–267) | 123 (31–269) | |
| Fruits | 43 (0–152) | 51 (0–192) | 46 (0–201) | 44 (0–176) | 28 (0–171) | 19 (0–147) | 25 (0–178) | 39 (0–230) | 53 (0–251) | |
| Non-leaf vegetables | 23 (4–102) | 35 (5–95) | 37 (5–130) | 38 (7–146) | 46 (8–145) | 49 (10–167) | 56 (11–182) | 60 (12–205) | 57 (9–195) | |
| Leaf vegetables | 10 (0–53) | 15 (0–65) | 18 (0–73) | 21 (0–87) | 25 (0–95) | 24 (0–88) | 29 (0–107) | 38 (1–123) | 39 (1–123) | |
| Vegetable products | 15 (0–65) | 18 (0–72) | 20 (0–77) | 27 (1–96) | 27 (1–100) | 24 (0–101) | 30 (1–114) | 35 (0–116) | 32 (0–112) | |

Table 2 Daily consumption rates of the different food groups (g) in women (median, 5th–95th percentiles)

| | Age group (years) | | | | | | | | | |
|---------------------|-------------------|--------------|---------------|---------------|---------------|----------------|----------------|----------------|---------------|--|
| | 4–6 (n=469) | 7–9 (n=382) | 10–12 (n=346) | 13–14 (n=291) | 15–18 (n=730) | 19–24 (n=1504) | 25–50 (n=5304) | 51–64 (n=2103) | > 64 (n=1179) | |
| Fish | 0 (0–29) | 0 (0–32) | 0 (0–38) | 0 (0–40) | 0 (0–45) | 0 (0–52) | 3 (0–55) | 11 (0–66) | 6 (0–64) | |
| Meat products | 72 (22–147) | 91 (30–188) | 111 (42–208) | 123 (49–263) | 117 (32–225) | 115 (25–240) | 128 (39–248) | 127 (50–243) | 121 (45–229) | |
| Eggs | 18 (1–47) | 19 (1–55) | 21 (3–52) | 21 (2–61) | 21 (2–57) | 22 (1–66) | 26 (2–68) | 25 (2–64) | 26 (2–65) | |
| Milk products | 292 (66–616) | 300 (48–656) | 279 (39–639) | 240 (37–715) | 197 (31–539) | 177 (35–511) | 144 (28–466) | 147 (29–437) | 158 (27–510) | |
| Pastries | 112 (52–223) | 154 (69–283) | 168 (73–310) | 175 (89–334) | 170 (64–328) | 167 (63–325) | 177 (66–339) | 198 (84–360) | 193 (89–366) | |
| Cereal products | 30 (7–95) | 32 (7–91) | 36 (8–94) | 33 (6–122) | 34 (5–107) | 35 (6–105) | 29 (4–90) | 22 (3–74) | 21 (4–74) | |
| Potato products | 58 (9–143) | 73 (19–170) | 80 (23–174) | 91 (15–197) | 80 (12–211) | 72 (7–188) | 83 (11–206) | 102 (23–226) | 106 (23–230) | |
| Fruits | 38 (0–164) | 50 (0–188) | 48 (0–200) | 47 (0–193) | 45 (0–199) | 34 (0–175) | 38 (0–190) | 59 (0–240) | 65 (0–252) | |
| Non-leaf vegetables | 25 (3–93) | 32 (3–122) | 35 (5–117) | 40 (7–131) | 44 (8–162) | 48 (9–163) | 55 (9–187) | 61 (10–194) | 53 (8–178) | |
| Leaf vegetables | 11 (0–48) | 17 (0–65) | 18 (0–71) | 22 (0–78) | 22 (0–89) | 22 (0–86) | 28 (0–104) | 36 (1–121) | 34 (0–120) | |
| Vegetable products | 13 (0–59) | 17 (0–68) | 19 (0–72) | 20 (0–83) | 19 (0–94) | 22 (0–91) | 27 (1–99) | 29 (0–107) | 28 (0–106) | |

Table 3 Significant correlations between the consumption of meat products and other food groups (Spearman correlation coefficients)

| | | Age group (years) | | | | | | | | |
|---------------------|---|-------------------|----------|---------|---------|---------|----------|----------|----------|----------|
| | | 4–6 | 7–9 | 10–12 | 13–14 | 15–18 | 19–24 | 25–50 | 51–64 | > 64 |
| Fish | M | | | | | | | –0.03* | –0.08** | –0.13*** |
| | F | | | | | | 0.06* | | –0.04* | |
| Eggs | M | 0.16** | 0.11* | | 0.14* | 0.11* | | 0.07*** | 0.07** | 0.08* |
| | F | 0.11* | | | | 0.09* | 0.16*** | 0.11*** | 0.12*** | 0.06* |
| Milk products | M | | | | | –0.10* | –0.12*** | –0.14*** | –0.10*** | –0.15*** |
| | F | | –0.11* | | | –0.10** | –0.16*** | –0.18*** | –0.11*** | –0.07* |
| Pastries | M | 0.20*** | 0.24*** | 0.29*** | 0.34*** | 0.23*** | 0.18*** | 0.19*** | 0.23*** | 0.24*** |
| | F | 0.19*** | 0.30*** | 0.28*** | 0.24*** | 0.21*** | 0.17*** | 0.18*** | 0.16*** | 0.24*** |
| Cereal products | M | –0.10* | –0.12* | –0.12* | | | | –0.04** | 0.08** | |
| | F | | –0.19*** | | | | | | 0.05* | |
| Potato products | M | 0.24*** | 0.18*** | 0.20*** | | 0.14** | 0.20*** | 0.18*** | 0.14*** | 0.18*** |
| | F | 0.31*** | 0.19*** | 0.17** | 0.18** | 0.33*** | 0.33*** | 0.25*** | 0.18*** | 0.12*** |
| Fruits | M | –0.11* | | | | | –0.10** | –0.05*** | | –0.07* |
| | F | | | | | | –0.12*** | –0.09*** | –0.06* | |
| Non-leaf vegetables | M | 0.11* | 0.11* | | 0.18** | 0.18*** | 0.11*** | 0.08** | 0.10*** | 0.10** |
| | F | 0.24*** | | 0.13* | 0.16** | 0.09* | | 0.08*** | 0.10*** | 0.12*** |
| Leaf vegetables | M | 0.16** | | 0.20*** | 0.16** | 0.12** | 0.15*** | 0.10*** | 0.14*** | 0.17*** |
| | F | 0.20*** | 0.18*** | | 0.15* | 0.14*** | 0.16*** | 0.12*** | 0.12*** | 0.13*** |
| Vegetable products | M | 0.18*** | 0.19*** | 0.14** | | 0.18*** | 0.16*** | 0.12*** | 0.15*** | 0.12*** |
| | F | 0.25*** | 0.19*** | 0.18** | 0.17** | 0.20*** | 0.16*** | 0.18*** | 0.16*** | 0.12*** |

M – male subjects; F – female subjects.
 *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

in Fig. 1, exemplified separately for males and females of age group 25 to 50 years. Sample parameters of location and spread of TRR values are provided in Table 5 for all age groups. Recalling the possible range of TRR from 0 to 11, we note that in all age groups the means and medians are about 5.5, which coincides with the centre of the TRR range. Also, the 95th percentiles differ only little and are all close to 7.6.

It is not surprising that no person in any age group

reached the theoretical TRR maximum of 11. The highest TRR values were all below this margin. This constitutes another strong indication that adding up the eleven 95th percentiles is likely to considerably overestimate the corresponding true quantile value of total food consumption. In order to quantify the extent of this bias we introduced a ‘factor of overestimation’, which measures the degree of overestimation with respect to the sample’s actual 95th percentile. Table 6 lists these factors, calculated

Table 4 Significant correlations between the consumption of cereal products and other food groups (Spearman correlation coefficients)

| | | Age group (years) | | | | | | | | |
|---------------------|---|-------------------|----------|----------|----------|----------|---------|----------|----------|----------|
| | | 4–6 | 7–9 | 10–12 | 13–14 | 15–18 | 19–24 | 25–50 | 51–64 | > 64 |
| Fish | M | | | | | | | –0.04* | | –0.08* |
| | F | 0.10* | | | | | | | –0.07** | –0.08** |
| Meat products | M | –0.10* | –0.12* | –0.12* | | | | –0.04** | 0.08** | |
| | F | | –0.19*** | | | | | | 0.05* | |
| Eggs | M | | | 0.13* | | 0.08* | 0.10** | 0.09*** | 0.10*** | 0.12*** |
| | F | | 0.10* | | | 0.11** | 0.07** | 0.10*** | 0.08*** | 0.12*** |
| Milk products | M | 0.13** | | | 0.23*** | 0.25*** | 0.18*** | 0.23*** | 0.22*** | 0.24*** |
| | F | –0.09* | 0.15** | 0.11* | 0.15* | 0.23*** | 0.14*** | 0.22*** | 0.20*** | 0.18*** |
| Pastries | M | –0.13** | –0.18*** | | | | –0.07* | | 0.05* | |
| | F | –0.12** | –0.14** | –0.16** | | | | 0.05*** | 0.07** | |
| Potato products | M | –0.18*** | –0.23*** | –0.20*** | –0.21*** | –0.20*** | –0.08** | –0.12*** | –0.12*** | –0.15*** |
| | F | –0.13** | –0.11* | –0.19** | –0.13* | –0.11** | –0.08** | –0.06*** | –0.08** | –0.17*** |
| Fruits | M | 0.11* | | | 0.19** | | 0.08** | 0.09** | 0.12*** | 0.07* |
| | F | | | | 0.01* | | 0.06* | 0.06** | 0.06** | 0.07* |
| Non-leaf vegetables | M | 0.11* | | | | 0.17*** | 0.13*** | 0.12*** | 0.10*** | 0.09** |
| | F | 0.14** | | | | 0.08* | 0.18*** | 0.10*** | 0.05* | 0.06* |
| Leaf vegetables | M | | 0.10* | | | | | 0.09*** | 0.07** | |
| | F | | 0.14** | | | | | 0.05*** | | |
| Vegetable products | M | | | | | 0.09* | | 0.03* | | |
| | F | | | | | 0.09* | 0.08** | 0.06*** | | |

M – male subjects; F – female subjects.
 *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Table 5 Location and spread of total relative ranks (TRR) of food consumption in different age groups

| Age group (years) | Male subjects | | | | | Female subjects | | | | |
|-------------------|---------------|------|--------|-----------------|---------|-----------------|------|--------|-----------------|---------|
| | Mean | SD | Median | 95th percentile | Maximum | Mean | SD | Median | 95th percentile | Maximum |
| 4–6 | 5.51 | 1.31 | 5.56 | 7.68 | 9.07 | 5.51 | 1.28 | 5.50 | 7.60 | 9.51 |
| 7–9 | 5.51 | 1.21 | 5.45 | 7.46 | 8.97 | 5.51 | 1.20 | 5.46 | 7.48 | 9.36 |
| 10–12 | 5.52 | 1.21 | 5.58 | 7.49 | 8.53 | 5.52 | 1.21 | 5.54 | 7.44 | 8.52 |
| 13–14 | 5.52 | 1.29 | 5.49 | 7.55 | 9.20 | 5.52 | 1.27 | 5.57 | 7.47 | 8.87 |
| 15–18 | 5.51 | 1.27 | 5.49 | 7.73 | 9.24 | 5.51 | 1.25 | 5.52 | 7.54 | 9.18 |
| 19–24 | 5.50 | 1.22 | 5.52 | 7.51 | 9.99 | 5.50 | 1.27 | 5.53 | 7.66 | 9.35 |
| 25–50 | 5.50 | 1.25 | 5.51 | 7.55 | 10.04 | 5.50 | 1.26 | 5.55 | 7.48 | 10.16 |
| 51–64 | 5.50 | 1.27 | 5.50 | 7.62 | 9.82 | 5.50 | 1.21 | 5.53 | 7.45 | 9.88 |
| >64 | 5.51 | 1.22 | 5.55 | 7.45 | 9.22 | 5.50 | 1.27 | 5.54 | 7.56 | 9.29 |

SD – standard deviation.

for each age group and sex. All factors show biases of between 34 and 53%. All estimated values (sums of food group percentiles) in Table 6 are substantially larger than the upper limits of the corresponding 99.9% confidence intervals.

Discussion

Only a little is currently known about interdependencies between the consumption rates of different food categories. We present here a first systematic empirical evaluation of the correlations predominant between consumption rates of major food groups. Based on data originating from a comprehensive survey, we determined for nine different age groups all statistically significant pairwise associations between 11 food categories. As a main result our analysis demonstrated the existence of positive as well as negative correlations between certain food groups. This means that the amounts of intake of different food groups cannot be regarded simply as positively correlated or as independent from one another. Instead the underlying correlation structure is complex, which is in line with and further supports the current trend in nutritional epidemiology of analysing dietary patterns rather than looking at individual nutrients or foods⁹.

In our analysis some correlations are found independently of age and sex: for instance, a consistently positive association of intake of meat products and pastries, potatoes and vegetables (Table 3). These correlations possibly reflect overall dietary habits that may be characteristic to the German population. However, another consistently negative association persists between the intake of potatoes and cereals (Table 4), indicating that both can substitute for each other as a carbohydrate source.

Other associations become prominent only in adolescence, e.g. the positive relationship between the consumption rates of cereal products and milk, eggs, fruits, and vegetables. Similarly, the association between meat products and milk products becomes significantly negative only in the age groups above 15 years. These correlations could at least in part be explained by alternative eating patterns, like more vegetarian diets, which often develop only later in life.

With respect to consumption differences between men and women, it is remarkable to note that a surprising concordance characterises the association patterns of men and women.

The negative correlation between consumption of meat products and fruits from age 19 onwards deserves special

Table 6 Factors of overestimation for each age group and sex

| Age group (years) | Male subjects | | | | | Female subjects | | | | |
|-------------------|---------------|---------------------------|------|--------------------------------|---------------------------|-----------------|---------------------------|------|--------------------------------|---------------------------|
| | <i>n</i> | 95% percentile (99.9% CI) | SD | Sum of food group percentiles* | Factor of overestimation† | <i>n</i> | 95% percentile (99.9% CI) | SD | Sum of food group percentiles* | Factor of overestimation† |
| 4–6 | 450 | 1349 (1261, 1436) | 26.4 | 1801 | 1.34 | 469 | 1126 (1050, 1202) | 22.9 | 1668 | 1.48 |
| 7–9 | 441 | 1503 (1410, 1596) | 28.1 | 2093 | 1.39 | 382 | 1373 (1282, 1463) | 27.4 | 1922 | 1.40 |
| 10–12 | 358 | 1578 (1469, 1688) | 33.0 | 2320 | 1.47 | 346 | 1441 (1339, 1544) | 30.9 | 1984 | 1.38 |
| 13–14 | 275 | 1854 (1694, 2014) | 48.1 | 2636 | 1.42 | 291 | 1524 (1395, 1653) | 38.8 | 2228 | 1.46 |
| 15–18 | 592 | 1941 (1824, 2057) | 35.3 | 2748 | 1.42 | 730 | 1427 (1353, 1500) | 22.3 | 2064 | 1.45 |
| 19–24 | 1182 | 1833 (1753, 1912) | 24.0 | 2641 | 1.44 | 1504 | 1323 (1274, 1372) | 14.9 | 2002 | 1.51 |
| 25–50 | 4974 | 1636 (1604, 1667) | 9.7 | 2451 | 1.50 | 5304 | 1345 (1319, 1371) | 7.8 | 2053 | 1.53 |
| 51–64 | 1699 | 1707 (1668, 1745) | 11.7 | 2485 | 1.46 | 2103 | 1410 (1369, 1450) | 12.3 | 2133 | 1.51 |
| >64 | 930 | 1651 (1581, 1722) | 21.4 | 2458 | 1.49 | 1179 | 1446 (1387, 1505) | 17.8 | 2195 | 1.52 |

CI – confidence interval; SD – standard deviation.

* Sum of the 95th percentiles of the 11 food groups.

† Calculated as the ratio of sum of food group percentiles over the sample's actual 95th percentile.

consideration. A high intake of meat products is usually associated with a high fat intake. Ursin *et al.*¹ found that individuals on a low-fat diet substitute certain carbohydrate-rich foods such as fruits and vegetables for fat. A high meat as well as a low fruit consumption are both discussed as risk factors for cardiovascular diseases and various cancers. Therefore, it should also be kept in mind that associations between the consumption rates of food groups that are independently associated with certain diseases should be considered as potential confounders in epidemiological studies investigating these relations.

In this context it may be also of interest that the negative correlation between meat and fish consumption, as observed in adults, indicates that individuals consuming high amounts of meat tend to eat less fish and vice versa. This may have implications for the intake of various nutrients and for the ingestion of radionuclides as well, especially since contamination of fish is an important pathway for many radionuclides¹⁰.

The negative correlation of meat products and the food groups fish, fruits and milk products in adults is largely in agreement with the results of Elmstahl *et al.*², who examined the relationships between meat and other food items in 11 648 Swedish subjects born between 1926 and 1945. After adjustment for energy intake the investigators divided meat ingestion into quantiles. They found that an increasing meat intake was associated with significantly decreasing intakes of poultry, fish, fruits, bread, cereals and cheese, in men as well as in women. A significantly negative correlation with milk ingestion was observed only in men.

The consumption of food is a major pathway for the internal radiation exposure of man. This pathway is considered when artificial or natural radionuclides are released into the environment with waste air or water by nuclear facilities, uranium mining and milling sites, or research and medical laboratories. The potential radiation exposure by food is therefore an important issue with regard to possible health risks. It is estimated by a weighted sum of the consumption rates of various food groups, with the concentrations of specific radionuclides in the corresponding food groups as weights. Consumption habits are therefore an important factor for estimation of the potential ingestion dose.

The proposed formula in the RPO Guideline, as described above, estimates the 95th percentile of total food consumption as a sum of the corresponding percentiles of the food groups. Implicitly, such an approach assumes very restrictive and unsubstantiated correlation structures between food groups; for realistic and conceivable situations negative correlations have essentially to be excluded. Normality, for example, requires perfect positive association between all food groups in order to make the estimator unbiased. Such correlation structure, however, is clearly contradicted by the presented data, where broad evidence for the

existence of quite a few negative correlations between intake rates of various important food groups is provided. Additionally, we were able to identify a substantial number of food group pairs that showed no significant correlation at all. Theoretical reasoning leaves little doubt about the direction of the bias (namely towards overestimation), but it cannot easily predict the size of the deviation from the real percentile value. In an attempt to quantify empirically the extent of bias attached to the proposed formula, we used a factor of overestimation and obtained, for the different age groups in our sample, values between 1.34 and 1.53, i.e. the overestimation of food intake ranged from 34 to 53% (Table 6). By asymptotics we determined for each sample percentile also a 99.9% confidence interval to demonstrate how far out to the right the RPO values are, e.g. for the smallest sample (boys aged 13 to 14 years) the RPO value is 2636, whereas the confidence interval's upper limit is 2014.

One has to keep in mind, however, that the radiation dose depends not only on the amount and composition of the food intake but also on the particular activity of the nuclides in them. Therefore, the real extent and direction of bias will depend strongly on the type of nuclear facility and its specific release of radionuclides, with respect to amount as well as composition.

Food originating from extensive agriculture, hunting and collecting (e.g. sheep meat and milk, goat meat and milk, game, berries, mushrooms) is often characterised by a relatively high accumulation of radionuclides. The consumption of such foodstuffs may lead to a higher exposure and, in general, diets with large fractions of plant food products lead to higher doses. Pröhl and Müller¹⁰ simulated different food consumption scenarios to investigate the impact of modifications in food intake on the estimated dose intake. They calculated annual effective dose commitment from the annual activity intake with food and respective ingestion dose factors, and found that the influence of dietary modification was relatively small if the medians of exposure were compared. Regarding higher percentiles, the differences were more pronounced. The most pronounced impact was found for ¹²⁹I, ¹³⁵Cs and ²²⁶Ra for a scenario with higher root uptake and enhanced consumption of lamb and sheep and goat's milk.

These results indicate that the dietary habits of the area considered must be studied in order to identify the site-related critical foodstuffs and that it might be useful to subdivide some food groups even further, e.g. the food group 'meat products' into the two subgroups (1) 'cattle and swine' and (2) 'goat and sheep'; or 'fruits' into the three of (1) 'berries', (2) 'drupe' and (3) 'pome fruit'. As is discussed currently, for the identification of the major dietary patterns of a particular sample, factor or cluster analyses might be alternative and complementary approaches^{9,11}.

We have shown that the formula proposed by the

Guideline to the RPO is likely to overestimate the true amount of radiation. It can be argued that, for safety reasons, such an approach is advisable in order to assume the worst case and therefore to start from the maximum possible ingestion dose and additionally consider appropriate safety margins.

However, the EURATOM Council Directive explicitly requests estimations to be as realistic as possible⁴. Furthermore, from a population health viewpoint, another important role is played by countermeasures. This was discussed in detail elsewhere¹². Interventions will depend on the estimated effective dose caused by a nuclear accident and the recommended dose limit. Some of these interventions, such as an excessive restriction to a single foodstuff, can be also of disadvantage to health. Therefore, after a large release of radionuclides, the critical situation encountered must be evaluated as realistically as possible, in order to take appropriate countermeasures and to avoid harmful over-reactions both by authorities and the public. The importance of this issue became evident from the results of a survey on the number of induced abortions following the Chernobyl accident¹³.

In conclusion, this analysis has shown that consumption rates of major food groups exhibit a complex correlation structure comprising not only positive, but also some negative correlations. These depend less on sex than on age, and probably on cultural, educational and many environmental factors. Therefore it is necessary to study these interdependencies in more detail since they may have far-reaching implications on various health issues, such as epidemiological risk studies or radiation exposure prediction. The correlations observed in our analysis are based on 7-day food records, which may not necessarily be representative of the usual dietary intake over a longer period of time. Nevertheless, the correlations presented demonstrate on an empirical basis that the assumptions of the risk assessment procedure for radiation have major flaws. In order to obtain a more reliable estimation of the ingestion of radionuclides via food, specific scenarios of nuclear events with various prototype dietary patterns need to be considered.

References

- 1 Ursin G, Ziegler RG, Subar AF, Graubard BI, Haile RW, Hoover R. Dietary patterns associated with a low-fat diet in the national health examination follow-up study: identification of potential confounders for epidemiologic analyses. *Am. J. Epidemiol.* 1993; **137**: 916–27.
- 2 Elmstahl S, Holmqvist O, Gullberg G, Johansson U, Berglund G. Dietary patterns in high and low consumers of meat in a Swedish cohort study. *Appetite* 1999; **32**: 1–13.
- 3 Kübler W, Anders HJ, Heeschen W, eds. *Ergebnisse der Nationalen Verzehrsstudie (1985–1988) über die Lebensmittel- und Nährstoffaufnahme in der Bundesrepublik Deutschland*. Vera-Schriftenreihe Band XI. Niederlehen: Wissenschaftlicher Fachverlag Dr. Fleck, 1994.
- 4 96/29/EURATOM Council Directive of 13 May 1996 laying down basic safety standards for the protection of the health workers and the general public against the dangers arising from ionising radiation, OJL-159 of 29/06/96.
- 5 Verordnung für die Umsetzung von EURATOM-Richtlinien zum Strahlenschutz, Bundesgesetzblatt Nr. 38, 20.07.2001, Artikel 1 “Verordnung über den Schutz vor Schäden durch ionisierende Strahlen (Strahlenschutzverordnung – StriSchV)”.
- 6 Bundesminister der Justiz, ed. Allgemeine Verwaltungsvorschrift zu § 45 Strahlenschutzverordnung: Ermittlung der Strahlenexposition durch die Ableitung radioaktiver Stoffe aus kerntechnischen Anlagen oder Einrichtungen vom 21. Februar 1990. *Bundesanzeiger* 1990; **42**(64a).
- 7 Adolf T. *Nationale Verzehrsstudie (NVS)*. *Verbundstudie Ernährungserhebung und Risikofaktorenanalytik (VERA)*. Dokumentation [Public Use File on CD-ROM].
- 8 Kübler W, Anders HJ, eds. *Lebensmittel- und Nährstoffaufnahme in der Bundesrepublik Deutschland. Ergänzungsband zum Ernährungsbericht 1992 auf der Basis der Nationalen Verzehrsstudie*. Vera-Schriftenreihe Band XII. Niederlehen: Wissenschaftlicher Fachverlag Dr. Fleck, 1994.
- 9 Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr. Opin. Lipidol.* 2002; **13**: 3–9.
- 10 Pröhl G, Müller H. Radiation exposure from radionuclides in ground water: an uncertainty analysis for selected exposure scenarios. *Radiat. Environ. Biophys.* 1996; **35**: 205–18.
- 11 Jacques PF, Tucker KL. Are dietary patterns useful for understanding the role of diet in chronic disease? *Am. J. Clin. Nutr.* 2001; **73**: 1–2.
- 12 Cigna AA. Health impacts of large releases of radionuclides. Interactions with human nutrition and other indices of population health. *Ciba Found. Symp.* 1997; **203**: 141–51; discussion 151–4.
- 13 Knudsen LB. Legally induced abortions in Denmark after Chernobyl. *Biomed. Pharmacother.* 1991; **45**: 229–31.
- 14 Mood AM, Graybill FA, Boes DC. *Introduction to the Theory of Statistics*, 3rd ed. New York: McGraw-Hill, 1974.

Appendix

The 95th percentile ξ_p of a continuous random variable X with strictly increasing distribution function is unique and defined as that value ξ_p such that $P(X \leq \xi_p) = 0.95$.

For a normally distributed (i.e. Gaussian) random variable X with mean μ and variance σ^2 , the 95th percentile is $\xi_p \approx \mu + 1.96\sigma$. The percentile ζ_p of a sum $Z = X + Y$ of two identically and jointly Gaussian distributed random variables X and Y is calculated by $\zeta_p \approx 2\mu + 1.96\sqrt{\sqrt{2} \times \sqrt{\sqrt{1+\rho}}\sigma}$, with ρ denoting the correlation coefficient. Only for $\rho = 1$ (perfect positive correlation) do we get $\zeta_p = 2 \times \xi_p$, i.e. only then is the percentile ζ_p of the sum identical to the sum of the percentiles. Otherwise percentile ζ_p is smaller, e.g. for independence (i.e. $\rho = 0$) we get $\zeta_p \approx 2\mu + 1.96\sqrt{\sqrt{2}}\sigma$.

Asymptotically, the 95th percentile x_p of a sample of size n is normally distributed with mean ξ_p and variance $0.0475 \times \left(f(\xi_p) \times \sqrt{\sqrt{n}} \right)^{-2}$, where $f(\xi)$ is the underlying probability density¹⁴. For a normally distributed population the standard error of ξ_p is therefore approximately $2.1 \times s_{\bar{x}}$, with $s_{\bar{x}}$ denoting the standard error of the sample's mean \bar{x} . Using this approximation allows one to calculate a confidence interval for x_p .