The Journal of Nutrition
Nutritional Epidemiology

Estimating Usual Food Intake Distributions by Using the Multiple Source Method in the EPIC-Potsdam Calibration Study[1–3]

Jennifer Haubrock,[4]* Ute Nöthlings,[5] Jean-Luc Volatier,[6] Arnold Dekkers,[7] Marga Ocké,[7] Ulrich Harttig,[4] Anne-Kathrin Illner,[4] Sven Knüppel,[4] Lene F. Andersen,[8] and Heiner Boeing[4] on behalf of the European Food Consumption Validation Consortium[9]

[4]Department of Epidemiology, German Institute of Human Nutrition, Nuthetal 14558, Germany; [5]Epidemiology Section, Institute for Experimental Medicine, University of Kiel, Kiel 24105, Germany; [6]Afssa French Food Safety Agency, Maisons-Alfort 94706, France; [7]National Institute for Public Health and the Environment, Bilthoven 3720, The Netherlands; and [8]Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, Oslo 0317, Norway

Abstract

Estimating usual food intake distributions from short-term quantitative measurements is critical when occasionally or rarely eaten food groups are considered. To overcome this challenge by statistical modeling, the Multiple Source Method (MSM) was developed in 2006. The MSM provides usual food intake distributions from individual short-term estimates by combining the probability and the amount of consumption with incorporation of covariates into the modeling part. Habitual consumption frequency information may be used in 2 ways: first, to distinguish true nonconsumers from occasional nonconsumers in short-term measurements and second, as a covariate in the statistical model. The MSM is therefore able to calculate estimates for occasional nonconsumers. External information on the proportion of nonconsumers of a food can also be handled by the MSM. As a proof-of-concept, we applied the MSM to a data set from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Calibration Study (2004) comprising 393 participants who completed two 24-h dietary recalls and one FFQ. Usual intake distributions were estimated for 38 food groups with a proportion of nonconsumers > 70% in the 24-h dietary recalls. The intake estimates derived by the MSM corresponded with the observed values such as the group mean. This study shows that the MSM is a useful and applicable statistical technique to estimate usual food intake distributions, if at least 2 repeated measurements per participant are available, even for food groups with a sizeable percentage of nonconsumers. J. Nutr. 141: 914–920, 2011.

JN THE JOURNAL OF NUTRITION

Introduction

The European Food Consumption Survey Method project has recommended to apply 24-h dietary recalls on at least 2 nonconsecutive days per participant (1) as the primary instrument for food consumption surveys (2,3) to account for intra-individual variation. The 24-h dietary recall is a short-term dietary assessment instrument that covers the consumption of foods during the day preceding the interview in great detail (4). Currently, the instrument is mostly administered with the help of computer programs, such as the European EPIC-Soft (5) or the US Automated Multiple Pass Method (6), which include probing questions and quality checks (7).

Recent advances in statistical methods (1,8–12) have provided statistical algorithms that derive usual intake distributions for populations based on multiple 24-h dietary recalls. The removal of intra-individual variability across consumption days is applied as a crucial processing step. These methods extended the work in the 1980s of Beaton et al. (13,14), who developed equations to estimate and remove the intra-individual part of the variation on a normal scale. However, the majority of these methods can be applied only to nutrients and foods that are consumed daily (1,9–12). In case of zero intakes, the application of these methods is challenging. Zero intakes occur in 24-h dietary recall data if food or food groups are only occasionally consumed. However, these nonconsumption days do not necessarily reflect

[1] Supported by the European Commission, 6th Framework Program (FOOD-CT-2006-022895). This document reflects only the author's views and the community is not liable for any use that may be made of the information contained therein.

[2] Author disclosures: J. Haubrock, U. Nöthlings, J-L. Volatier, A. Dekkers, M. Ocké, U. Harttig, A-K. Illner, S. Knüppel, L. F. Andersen, and H. Boeing, no conflicts of interest.

[3] Supplemental Tables 1 and 2 as well as the detailed statistical background of the Multiple Source Method are available from the "Online Supporting Material" link in the online posting of the article and from the same link in the online table of contents at jn.nutrition.org.

[9] Members of the European Food Consumption Validation Consortium include H. Boeing, J. Haubrock, U. Harttig, E. Trolle, P. Amiano, S. Henauw, M. Maeyer, M. Bellemans, I. Huybrechts, W. Keyzer, M. Feinberg, N. Slimani, C. Casagrande, G. Nicolas, C. Leclerqs, C. Donne, E. Boer, H. Brants, M.C. Ocke, C. Rossum, C.W. Hooven, Z. Etemad, J. Laan, P. Veer, J. Vries, A. Geelen, P. Hulshof, S. Crispim, O. Souverein, L.F. Andersen, I.T. Lillegaard, J. Ruprich, I. Rehurkova, M. Dofkova, J.L. Volatier, L. Lafay, S. Lioret, A. Kaic-Rak, J.P. Cvetkovic, E. Purse.

* To whom the correspondence should be addressed. E-mail: Jennifer.Haubrock@dife.de.

Manuscript received December 21, 2010. Initial review completed March 13, 2010. Revision accepted January 31, 2011.
First published online March 23, 2011; doi:10.3945/jn.109.120394.

true nonconsumption. In fact, usual nonconsumption remains uncertain if no additional information on habitual food intake is available.

In following studies, 24-h dietary recall information was supported with additional information on frequency of food consumption to overcome this limitation. Nusser et al. (8) proposed to calculate the usual intake distribution focusing on participants who declared themselves as consumers. Recently, a group of the National Cancer Institute (NCI)[10] recommended using the full scale of frequency information, including null consumption as covariate information for estimation of the usual dietary intake distribution (15). In comparison with the Nusser method, the concept of the NCI group is more progressive, because it makes use of the information about nonconsumption as an integrated part of the overall estimation process (15).

In correspondence with the strategy of the NCI group (15), the Multiple Source Method (MSM) was developed to estimate usual intake distributions within the framework of the European Food Consumption Validation project.

In relation to the estimation concept, the MSM involves similar steps to those proposed by the NCI group to derive usual intake distributions by estimating the probability and the amount of consumption and combining both estimations. Both the MSM and NCI method include covariates into the modeling parts (15). Furthermore, both methods can be applied to nutrients, foods, or occasionally or rarely consumed foods if at least 2 repeated measurements for some participants can be provided. However, both methods differ in terms of the modeling part, handling nonconsumers, inclusion of external information within the estimation process, coping with correlations, and providing the method to potential users.

In this paper, we present the application of the MSM to a data set of the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Calibration Study along with 2 simulation studies. The theoretical background of the MSM as well as the back transformation equations and the results from the simulation studies (**Supplemental Tables 1** and **2**) are available as Online Supporting Material. The specific aspects of the MSM are discussed on the basis of these results.

Participants and Methods

Study population and dietary assessment. In the area of Potsdam, Germany, 27,548 participants aged between 35 and 65 y were recruited from the general population between 1994 and 1998 for the EPIC-Potsdam Study (16). Baseline assessments included diet and lifestyle questionnaires and measurements of weight and height among other characteristics. During active follow-up, dietary intake of the preceding year was assessed by a simplified FFQ that contained 102 food items. The development of this FFQ with its implementation of portion size information has been previously described in detail (17). The self-administered FFQ inquired only about the frequency of consumption using a closed-ended format of discrete categories that ranged from "never", "one time per month" to "11 times per day or more frequent" and can, therefore, be considered as being a food propensity questionnaire according to the definition of Subar et al. (18).

In 2004, a study was conducted to validate the simplified FFQ using repeated 24-h dietary recalls (17), the so-called EPIC-Potsdam Calibration Study. A random sample of 230 men and 230 women was drawn from the original sample to provide 2 unannounced recalls within 1 y and to complete the simplified FFQ at the end of the year. The sampling days were randomized over the year preceding the administration of the FFQ, thus reflecting consumption habits during the time period also covered by the FFQ. For collecting 24-h dietary recalls, we used the standardized EPIC-SOFT Program (5,19) administered by telephone interviews. To better estimate portion sizes during the interview, an adapted version of the EPIC-SOFT picture book was mailed to the participants in advance.

Four hundred of the 460 initially invited participants (87%) provided 24-h dietary recalls. After exclusion of 1 participant with no second 24-h dietary recall and 6 participants with no FFQ, a complete set of two 24-h dietary recalls and one simplified FFQ was available for 393 study participants (197 men and 196 women), forming the data basis for this analysis. The mean time between recalls was 178.4 d and the mean time between the last recall and the FFQ was 104.5 d.

For a description of the study population, we used the variables education, weight, height, and smoking status of the EPIC-Potsdam lifestyle questionnaire at baseline. The 24-h dietary recall consumption data of the EPIC-Potsdam Calibration Study and consumption frequency information of the simplified EPIC-Potsdam FFQ were used for statistical analysis (17).

Statistical methods. Due to the small sample size, a separate statistical analysis for men and women was not performed.

For each individual, all reported item frequencies from the simplified FFQ were converted to mean frequencies per day. For example, a report of 1 time/wk was converted to one-seventh times per day. These frequencies were then summed into reported frequencies at food group level of 39 food groups, consistent with previous studies (20). Dietary intake information provided by the 24-h dietary recall was also collapsed into the same 39 food groups and absolute intakes in g/d per food group were computed. Because no participant consumed foods that belong to the food group "snacks" on both recalled days, the following results consider 38 of 39 food groups. Because the MSM needs repeated measurements on at least 1 individual to estimate intra-individual variance, this group was excluded from further analyses.

Additional food frequency information was used in 2 different ways within the MSM. Frequency information from the FFQ allowed us to separate occasional nonconsumers from those who were true nonconsumers. Participants who indicated zero consumption of a food or food group on the FFQ were classified as nonconsumers if they additionally did not report consumption of the food group in either 24-h dietary recall and were defined as true nonconsumers. For true nonconsumers, the probability of consumption as well as the food intake on consumption days was set to zero. Those who were not true nonconsumers but had no consumption in the 24-h dietary recalls were considered for 1the estimation procedure. Their consumption amount was estimated through simulation taking covariates into account. In this study, frequency information was also used as covariate within the modeling parts of the MSM.

The applied MSM comprised 3 steps based on at least 2 repeated short-term measurements. First, for each individual in the study sample, the probability of consumption of a food group on a randomly selected day was calculated. Second, the usual amount of food group intake on reported consumption days was estimated, and finally, the usual overall food group intakes were calculated by multiplying probability of consumption of a food group with usual amount of food group intake on consumption days.

In the course of this study, occurrence variables were defined for each 24-h dietary recall day in a first step, with a value of 1 indicating consumption on a 24-h dietary recall day and zero indicating nonconsumption. A logistic regression was applied to model the occurrence variable as a function of a list of covariates (sex, age, sex and age, and FFQ frequency), which were assumed to be predictive for consumption of a food group. The probability that a participant consumed a specific food on one day was estimated together with corresponding model residuals. These residuals were transformed to real numbers, inter- and intra-individual variances were estimated, intra-individual variances were excluded, and subsequently shrunken residuals were back transformed to the original scale.

The second step of the MSM was restricted to the observed food group intake data in 24-h dietary recalls. On these intake data, a linear regression model was applied. Similar to the first step, the consumption was modeled as a function of a consistent set of covariates that were

[10] Abbreviations used: EPIC, European Prospective Investigation into Cancer and Nutrition; MSM, Multiple Source Method; NCI, National Cancer Institute.

assumed to be predictive for the observed amount of food intake of each individual and model-based predicted values were obtained. The corresponding residuals of the linear regression model were transformed to normality by a 2-parameter Box-Cox transformation family (10). In case a participant reported no consumption of a food group on any of the two 24-h dietary recalls but reported consumption according to the FFQ, usual intake on consumption days was estimated in the transformed scale taking covariate information into account.

In the transformed scale, we assumed the classical measurement error model. Using the transformed residuals, we estimated the inter-individual variance. The resulting quantities were then back transformed to the original scale and added to the model prediction to obtain the usual food intake on consumption days.

In the last step, the distribution of usual food intake for the study population was estimated by multiplication of the results of steps 1 and 2. More in-depth statistical information of the MSM is provided as Online Supporting Material. To describe the usual food intake distributions, percentiles (5th–95th) and 4 moments, mean ± SD, Kurtosis, and Skewness, were reported.

Additionally, 2 simulation studies were conducted to compare estimates obtained by different back transformation equations. Detailed simulation results as well as the applied equations for back transformation are available as OSM.

The aim of the first simulation study was to estimate usual intake distribution on consumption days. The accuracy of the estimated usual intake distribution was highly dependent on the back transformation equation after a Box-Cox transformation was applied in step 2 of the MSM. For data back transformation, we applied Eq. 13 of Hoffmann et al. (10) and Eq. 14 of Dodd et al. (21) (OSM). To compare the accuracy of Eq. 13 and 14, we assumed that the original data must be transformed to normality by using a 2-parameter Box-Cox transformation function with τ = 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9, or 1/10 as power transformation parameter, which correspond to λ = 2–10 (data not shown for $\tau = {}^1/_2 – {}^1/_9$). According to the 2-parameter Box-Cox transformation, function λ has to be a positive integer, otherwise the advantage of an exact back transformation does not hold. In the transformed scale, we generated data for 100,000 d for each of the 1000 participants modeling both inter-individual and intra-individual variation by normal distributions with mean 0 and variance 1. To exclude intra-individual variation and to receive corrected distribution parameters, the individual means in the normal scale were subsequently shifted to the sample mean and these shrunken values were back transformed to the original scale by applying either of Eq. 13 or 14. Afterwards, the distribution was compared with the true usual intake distribution, which can be generated by applying the inverse Box-Cox function on the daily data and forming individual means over the 100,000 d.

In a second simulation study, we explored the accuracy of the MSM to estimate consumption probabilities. For this purpose, we first generated consumption P for 1,000,000 individuals assuming a uniform distribution over the interval (0, P^*), with $P^* \leq 1$. Second, we simulated the event of consumption on each of 2 sampling days based on a Bernoulli distribution with parameter P. From these 2 observations, we finally estimated the consumption probability for each participant using the MSM without covariate information.

All analyses were performed using SAS software (version 9.1, SAS Institute).

Results

The mean age of the participants was 57 y (**Table 1**). Food groups with a proportion of nonconsumers > 70% in the 24-h dietary recalls were legumes, other fruits, nuts, other alcoholic beverages, breakfast cereals, and spirits (**Table 2**). In contrast, the highest proportions (>50%) of true nonconsumers were observed for the food group breakfast cereals and spirits.

We estimated the usual food intake distributions with the MSM for all but 1 food group (**Table 3**). The usual intake distribution for the food group "snacks" could not be estimated, because there was no study participant that consumed food belonging to this food group on both recall days. Therefore, the estimation of intra-individual variation was not possible and the MSM could not be applied. Intakes of the remaining food groups were described by percentiles and corresponding 4 moments (mean ± SD, Kurtosis, and Skewness).

TABLE 1 Participant characteristics in the EPIC-Potsdam Calibration Study, 2004[1]

Characteristics	
Age, *y*	57 ± 8.9
BMI, *kg/m²*	26.9 ± 4.4
Education, *%*	
No vocational/vocational training	35
Technical school	24
University degree	40
Smoking status, *%*	
Never	43
Former	42
Current	7

[1] Values are percent or means ± SD, n = 393.

All proportions of true nonconsumers as well as empirically derived mean were correctly reflected in the estimated usual food intake distributions. For example, after application of the MSM, the food group "breakfast cereals" (true nonconsumer proportion of 50%) appeared not to be consumed by one-half of the study population as reflected in the percentiles. Other rarely consumed food groups like nuts and other fruits with much lower true nonconsumer proportions showed estimated zero intakes up to the 25th percentile.

The resulting percentiles and distributional parameters from the first simulation study to estimate usual intake distributions on consumption days (with $\tau = {}^1/_{10}$) are described in the top of Supplemental Table 1. Comparing these distributions with the true usual intake distribution revealed that the NCI back transformation equation underestimates usual intake, because it uses only 2 terms, whereas the MSM back transformation is accurate. By increasing the intra-individual variance from 1 to 5, the bias of the NCI approximate back transformation equation was more apparent (Supplemental Table 1). For a variance ratio of 5, all percentiles were underestimated by more than 25%. The true population mean and SD of usual intake were underestimated by 34 and 24%, respectively.

As long as the inverse of the power parameter was >3, the approximate back transformation equation generated biased results. All percentiles and the SD were underestimated by the approximate equation and the degree of underestimation increased with decreasing power parameter. In contrast to the NCI equation, the exact solution, Eq. 13, always reproduced the true usual intake distribution, for $1/\tau=\lambda$ with λ a positive integer (λ= 2–10).

The results from the second simulation study that aimed to explore the accuracy of the MSM to estimate consumption probabilities are given in Supplemental Table 2. The mean and the SD of the estimated consumptions probabilities can be compared with the true values. Obviously, the difference between estimated and true parameters was always small and did not depend on the range of the simulated consumption probabilities. The estimation methods were unbiased and consistent.

Discussion

This study shows that the MSM could be successfully applied in estimating usual diet from short-term data accrued in the EPIC-

TABLE 2 Patterns of observed nonconsumption and consumption in 24-h dietary recalls compared with FFQ in the EPIC-Potsdam Calibration Study, 2004[1]

Food groups	True nonconsumer	Nonconsumer in 24-h dietary recalls
	%	
Bread	0	0
Milk and dairy products	0	5
Sugar and confectionary	0	7
Fruiting and root vegetables	0	7
Processed meat	0	8
Fresh fruits	0	9
Cheese	0	12
Other vegetables	0	16
Condiments	0	22
Margarines	0	23
Butter and other animal fat	0	25
Sauces	0	25
Potatoes	0	27
Vegetable oils	0	37
Red meat	0	46
Eggs	0	52
Soups	0	52
Pasta, rice	0	61
Leafy vegetables	0	62
Other cereals	0	63
Cabbages	0	65
Poultry	0	74
Legumes	0	91
Water	1	7
Cakes, cookies	1	24
Fruit and vegetable juices	1	30
Soft drinks	1	76
Other fruits	1	86
Fish	2	61
Coffee	3	7
Tea	4	29
Desserts	5	74
Nuts	7	82
Wine	10	63
Other alcoholic beverages	23	89
Beer	26	66
Breakfast cereals	50	89
Spirits	51	92

[1] n = 393.

Potsdam Calibration Study. The intake estimates derived by the MSM corresponded with the observed values such as the percentage of true nonconsumers and the group mean of the 24-h dietary recalls. For 38 food groups, we obtained usual food intake distributions described by percentiles and 4 moments (mean ± SD, Kurtosis, and Skewness).

To assess the significance of the MSM method, its characteristics as well as similarities to the NCI method (15) are discussed in the following.

In terms of the 2-part estimation concept and the inclusion of covariates in the modeling parts, the MSM uses similar steps as those proposed by the NCI group (15,22) to derive usual intake distributions. A precondition for including covariates within the MSM model to improve the estimation of consumption probability and intake amount is that covariate information is available for each participant in the study.

Furthermore, both methods can be applied to nutrients and foods even if they are consumed occasionally if at least 2 repeated measurements per participant are provided.

The MSM as well as the NCI method (15) account for data characteristics described by Dodd et al. (21). First, the MSM accounts for the typical spike at zero due to nonconsumption on recalled days by multiplying the probability of consuming a food and the amount consumed on a consumption day. Second, the MSM addresses challenges by transforming skewed intake data distributions to normality. Third, methodical challenges such as distinguishing intra- and inter-individual variances and incorporating covariate information are addressed by the MSM.

Consequently, when estimating usual food intake distributions, both methods should be of similar efficiency with λ as a positive integer. However, our results from the simulation study (OSM) and a comparison study of Souverein et al. (23) revealed differences caused by different back transformations. In our simulation study, the exact back transformation used in the MSM showed better efficiency compared with the back transformation of the NCI method. Our simulation results suggest that the NCI method will lead to similar results once the back transformation is corrected. Therefore, it could be considered to use the back transformation applied in this study for restricted λ or by simulations as described in (11) and (A. Dekkers, M. Ocke, W. Slob, unpublished data) with the NCI method to further improve its results. Accordingly, the NCI method was recently updated (22). The authors advise the use of a different back transformation when the intra-individual variation is much larger than the inter-individual variation and in cases where the data are strongly skewed.

The MSM as well as the NCI method primarily utilize information obtained by a repeated short-term instrument. These can be repeated 24-h dietary recalls but also 24-h protocols or records with at least 2 dietary assessments per participant. In the European Food Consumption Validation proposal as well as in this study, the instrument was a computerized 24-h dietary recall. To apply the MSM, the requirement for repeated short-term measurements is their statistical independence. Therefore, randomly selected consumption occasions are needed. It can be assumed that food intakes of consecutive days are highly correlated such that high intake on one day is followed by low intake on the next day (24).

In terms of possible correlations between probability of consumption and consumption amount, we assumed that the MSM lead to nearly unbiased estimates if the sampled recall days were statistically independent of each other, which was usually the case when the days were randomly selected (no consecutive days) and when frequency was used as a covariate. The study design of the EPIC-Potsdam Calibration Study accounted for this precondition, because both participants and sampling days were randomly selected. Sampling days were collected over a complete calendar year, ensuring that the different days of a week were drawn with similar frequencies and that the 4 seasons were equally represented. Thus, no preliminary data adjustment for the day of the week or season was necessary. In comparison, the NCI method accounts for correlation by simultaneously estimating the model parameters and adding an extra model parameter (15,22).

Compared with the NCI method, the MSM is different in modeling parts (transforming residuals only) and the character of transformation (restricted 2-parameter Box-Cox transformation) and back transformation (exact integral solution for certain τ values). The general concept of the NCI is a 2-part mixed-effects regression model to simulate data for estimating

TABLE 3 Usual food intake distributions estimated using the MSM for 38 food groups in the EPIC-Potsdam Calibration Study, 2004[1]

Food group (% true nonconsumers)	Percentile 5	10	25	50	75	90	95	Moments: Estimated mean ± SD	Observed mean	Kurtosis	Skewness
						g/d					
Bread (0)	82	91	112	144	184	226	270	154 ± 59	154	2.29	1.17
Milk and dairy products (0)	39	55	86	150	242	341	436	183 ± 128	183	3.61	1.55
Sugar and confectionary (0)	11	14	23	35	48	68	82	39 ± 25	40	5.84	1.86
Fruiting and root vegetables (0)	44	54	71	94	125	154	178	103 ± 46	104	3.94	1.49
Processed meat (0)	24	29	37	53	79	97	113	59 ± 27	60	−0.52	0.58
Fresh fruits (0)	67	102	145	239	350	469	544	264 ± 148	265	0.26	0.76
Cheese (0)	10	14	22	35	52	68	80	39 ± 22	39	1.24	0.95
Other vegetables (0)	16	19	24	31	40	50	55	33 ± 12	35	−0.06	0.52
Condiments (0)	1	1	2	4	6	8	10	4 ± 4	4	3.82	1.55
Margarines (0)	1	3	5	13	23	36	42	16 ± 14	17	1.65	1.20
Butter and other animal fat (0)	1	3	5	10	18	30	38	14 ± 13	14	5.43	1.93
Sauces (0)	13	15	21	32	48	65	73	37 ± 19	37	0.90	0.90
Potatoes (0)	34	41	53	75	98	128	144	79 ± 33	79	0.79	0.81
Vegetable oils (0)	1	2	3	4	6	8	8	5 ± 2	5	−0.59	0.43
Red meat (0)	8	10	14	36	62	83	99	41 ± 29	41	−0.47	0.66
Eggs (0)	7	8	10	15	20	26	29	16 ± 8	16	3.11	1.39
Soups (0)	10	13	22	36	58	81	95	43 ± 28	43	2.28	1.28
Pasta, rice (0)	11	14	22	30	43	60	71	35 ± 19	35	6.41	1.70
Leafy vegetables (0)	1	3	5	10	17	28	40	14 ± 13	14	20.62	3.30
Other cereals (0)	0	1	2	5	10	16	21	7 ± 7	7	4.20	1.83
Cabbages (0)	5	7	11	18	29	41	52	22 ± 15	21	3.96	1.64
Poultry (0)	3	5	8	12	16	23	26	13 ± 8	13	4.63	1.58
Legumes (0)	0	0	0	1	2	6	31	5 ± 14	5	22.77	4.51
Water (1)	136	170	414	767	1121	1582	1868	830 ± 537	835	0.59	0.85
Cakes, cookies (1)	22	26	39	57	82	102	120	62 ± 32	62	1.33	0.86
Fruit and vegetable juices (1)	19	44	72	128	267	376	523	191 ± 181	197	8.58	2.39
Soft drinks (1)	1	3	7	13	55	169	307	74 ± 191	76	29.99	5.12
Other fruits (1)	0	0	0	0	1	15	31	5 ± 16	5	17.59	4.11
Fish (2)	4	7	11	19	44	61	71	28 ± 23	28	1.13	1.20
Coffee (3)	64	125	283	424	565	712	801	430 ± 227	434	1.39	0.50
Tea (4)	2	30	99	274	560	824	1157	385 ± 381	391	4.82	1.86
Desserts (5)	0	3	6	11	23	47	61	19 ± 23	19	14.55	3.07
Nuts (7)	0	0	1	1	2	10	15	3 ± 7	3	28.15	4.54
Wine (10)	0	0	8	19	115	199	246	71 ± 94	74	4.93	1.98
Other alcoholic beverages (23)	0	0	0	3	6	15	29	7 ± 14	7	27.45	4.68
Beer (26)	0	0	0	24	307	640	878	190 ± 290	203	2.55	1.76
Breakfast cereals (50)	0	0	0	0	1	12	20	3 ± 8	3	13.83	3.49
Spirits (51)	0	0	0	0	1	3	8	2 ± 6	2	54.94	6.55

[1] $n = 393$.

the usual intake distribution. The amount part of the model is transformed to normality conditionally on covariates using a one-parameter Box-Cox transformation with positive real-valued power parameter. In contrast to the NCI method (15,22), the MSM transforms only residuals that account for previously included model covariates. This technical difference is due to the fact that the intra-individual variance is reflected in the residuals.

The MSM computes usual intake distributions based on individual estimates in contrast to other existing methods (9,11,15), which calculate usual intake on the basis of simulated data. The ability to estimate usual food intake distributions on the basis of simulated data is limited, because it does not assign estimates of intake data to individuals nor does it consider the individual means in the estimation procedure. However, within the MSM, percentiles and moments of the population are derived by estimating individual usual intakes. This approach leads to estimated intake distributions of the population, which are closely linked to the sample size and to the empirical density function.

Furthermore, the MSM is specific in terms of handling usual nonconsumer of a food or a food group in the statistical model. A true nonconsumer remains a true nonconsumer and is assigned with a usual intake of zero. In the NCI method (15,22), the model does not incorporate never-consumers. However, the MSM calculates usual intake estimates in consideration of covariates for occasional nonconsumers too. This is in case a participant does not report consumption of a food within the repeated 24-h dietary recalls and if an auxiliary instrument (i.e. FFQ) is available. As we could see in our empirical results, a FFQ or another long-term measurement of intake greatly improves the estimates, especially for rarely or occasionally consumed foods. However, the MSM can still work without such information, because additional FFQ information is not always available. A major strength of the MSM is its ability to combine dietary intake data such as 24-h dietary recalls with supporting data on consumption frequency of other external sources, which is particularly important for the estimation of rarely consumed food distributions in some instances.

The implementation of the MSM is publicly available through a Web-based program (25) that can be accessed at the Web site (26). The Web-based approach, requiring only a modern Web browser as analysis tool, distinguishes MSM from other methods that provide either a stand-alone program as PC-Side (8,27) or a SAS macro for the NCI method (15,22,28). For analyzing data, researchers are not forced to install software or to pay any software license fees.

In this study, we did not provide confidence limits for our percentiles and moments. This option should be considered in further improving the MSM. The population estimates can, furthermore, be improved by incorporating reference biomarkers (29). Nevertheless, the estimated usual food intake distribution should be carefully interpreted, particularly in the case of a very low number of participants in the 24-h dietary recalls who repeatedly consumed from a specific food group.

In summary, the application of the MSM on food intake data was successful, even for rarely consumed food groups with high proportions of true nonconsumers. We described the food group intake of our population with a statistical function. Such data are rarely available but are desired by nutritionist and food risk assessors. Our results suggest that future dietary assessment in a survey should preferably include at least 2 nonconsecutive days of short-term assessment. To accurately and precisely estimate usual intake distributions and the proportion of true nonconsumers for rarely consumed foods within a population, we highly recommend repeated short-term measurement information with frequency information such as a FFQ. If no additional consumption frequency information is available, the MSM therefore provides the unique possibility to include external information on the proportion of the true nonconsumers in a population within the estimation process.

Acknowledgements

The MSM was developed by our colleague Dr. Kurt Hoffmann, who unexpectedly passed away in 2007. We acknowledge his substantive contribution to this article. We thank our colleagues, Wolfgang Bernigau, Ellen Kohlsdorf, Elektra Polychronidou, and Herbert Piechot, who helped with the data collection and management. J.H. carried out the statistical analyses, prepared the tables and figures, and wrote the paper, taking into account comments from all coauthors; H.B. was the overall coordinator of this project; U.N., J-L.V., H.B., M.O., U.H., A-K.I., S.K., and L.F.A. were members of the writing group. They gave input on the drafting of the manuscript and interpretation of results. A.D. and S.K. gave input on the statistical analyses. All authors read and approved the final manuscript.

Literature Cited

1. Biró G, Hulshof K, Ovesen L, Amorim Cruz J. Selection of methodology to assess food intake. Eur J Clin Nutr. 2002;56 Suppl 2:S25–32.
2. Brussaard J, Lowik M, Steingrimsdottir L, Moller A, Kearney J, De Henauw S, Becker W. An European food consumption survey method-conclusions and recommendations. Eur J Clin Nutr. 2002;56 Suppl 2: S89–94.
3. Dwyer JT. Dietary assessment. In: Shils ME, Olson JA, Shike M, editors. Modern nutrition in health and disease. Philadelphia: Lea & Febiger; 1994. p. 842–60.
4. Thompson FE, Subar AF. Dietary assessment methodology. In: Coulston A, Boushey C, editors. Nutrition in the prevention and treatment of disease. Burlington (MA): Elsevier Academic Press; 2008. p. 3–22.
5. Slimani N, Ferrari P, Ocke M, Welch A, Boeing H, Liere M, Pala V, Amiano P, Lagiou A, et al. Standardization of the 24-hour diet recall calibration method used in the european prospective investigation into cancer and nutrition (EPIC): general concepts and preliminary results. Eur J Clin Nutr. 2000;54:900–17.
6. Blanton CA, Moshfegh AJ, Baer DJ, Kretsch MJ. The USDA Automated Multiple-Pass Method accurately estimates group total energy and nutrient intake. J Nutr. 2006;136:2594–9.
7. Subcommittee on Criteria for Dietary Evaluation, Coordinating Committee on Evaluation of Food Consumption Surveys, Food and Nutrition Board, National Research Council Authoring Organizations. Nutrient Adequacy NRC. Assessment using food consumption surveys. Washington, DC: National Academies Press; 1986. Available from: http://www.nap.edu/catalog.php?record_id=618.
8. Nusser SM, Carriquiry AL, Dodd KW, Fuller WA. A semiparametric transformation approach to estimating usual daily intake distributions. J Am Stat Assoc. 1996;91:1440–9.
9. Nusser SM, Fuller WA, Guenther PM. Adjusting for measurement error and nonnormality in 24-hour food intake data. In: Estimating usual dietary intake distributions. New York: Wiley; 1997.
10. Hoffmann K, Boeing H, Dufour A, Volatier JL, Telman J, Virtanen M, Becker W, De Henauw S, EFCOSUM Group. Estimating the distribution of usual dietary intake by short-term measurements. Eur J Clin Nutr. 2002;56 Suppl:S53–62.
11. Waijers PM, Dekkers AL, Boer JM, Boshuizen HC, van Rossum CT. The potential of AGE MODE, an age-dependent model, to estimate usual intakes and prevalences of inadequate intakes in a population. J Nutr. 2006;136:2916–20.
12. Guenther PM, Kott PS, Carriquiry AL. Development of an approach for estimating usual nutrient intake distributions at the population level. J Nutr. 1997;127:1106–12.
13. Beaton GH, Milner J, Corey P, McGuire V, Cousins M, Stewart E, de Ramos M, Hewitt D, Grambsch P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Am J Clin Nutr. 1979;32:2546–59.
14. Beaton GH, Milner J, McGuire V, Feather TE, Little JA. Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins, and minerals. Am J Clin Nutr. 1983;37:986–95.
15. Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ, Kipnis V. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. J Am Diet Assoc. 2006;106:1575–87.
16. Boeing H, Wahrendorf J, Becker N. EPIC-Germany-A source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. Ann Nutr Metab. 1999;43:195–204.
17. Nöthlings U, Hoffmann K, Bergmann MM, Boeing H. Fitting portion sizes in a self- administered food frequency questionnaire. J Nutr. 2007;137:2781–6.
18. Subar AF, Dodd KW, Guenther PM, Kipnis V, Midthune D, McDowell M, Tooze JA, Freedman LS, Krebs-Smith SM. The food propensity questionnaire: concept, development, and validation for use as a covariate in a model to estimate usual food intake. J Am Diet Assoc. 2006;106:1556–63.
19. Slimani N, Valsta L. Perspectives of using the EPIC-SOFT programme in the context of pan-European nutritional monitoring surveys: methodological and practical implications. Eur J Clin Nutr. 2002;56 Suppl: S63–74.
20. Schulz M, Hoffmann K, Weikert C, Nothlings U, Schulze MB, Boeing H. Identification of a dietary pattern characterized by high-fat food choices associated with increased risk of breast cancer: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Br J Nutr. 2008;100:942–6.
21. Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, Midthune D, Tooze JA, Krebs-Smith SM. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. J Am Diet Assoc. 2006;106:1640–50.
22. Tooze JA, Kipnis V, Buckmann DW, Carroll RJ, Freedman LS, Guenther PM, Krebs-Smith SM, Subar AF, Dodd KW. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. Stat Med. 2010;29:2857–68.

23. Souverein OW, Dekkers A, Geelen A, Haubrock J, deVries J, Ocke M, Harttig U, Boeing H, van'tVeer P. Comparing four methods to estimate usual intake distributions. Eur J Clin Nutr. In press 2011.
24. Hartman AM, Brown CC, Palmgren J, Pietinen P, Verkasalo M, Myer D, Virtamo J. Variability in nutrient and food intakes among older middle-aged men. Implications for design of epidemiologic and validation studies using food recording. Am J Epidemiol. 1990;132:999–1012.
25. Harttig U, Haubrock J, Knüppel S, Boeing H. The MSM program: Web-based statistics package for estimating usual dietary intake using the Multiple Source Method. Eur J Clin Nutr. In press 2011.
26. The German Institute of Human Nutrition [cited 2011 Jan 28]. Available from: https://nugo.dife.de/msm.
27. Iowa State University [cited 2011 Jan 28]. Available from: http://cssm.iastate.edu/software/side.html.
28. National Cancer Institute [cited 2011 Jan 28]. Available from: http://riskfactor.cancer.gov/diet/usualintakes/macros.html.
29. Freedman LS, Midthune D, Carroll RJ, Krebs-Smith S, Subar AF, Troiano RP, Dodd K, Schatzkin A, Bingham SA, et al. Adjustments to improve the estimation of usual dietary intake distributions in the population. J Nutr. 2004;134:1836–43.