SPADE, a New Statistical Program to Estimate Habitual Dietary Intake from Multiple Food Sources and Dietary Supplements1–3

Arnold LM Dekkers,4,5* Janneke Verkaik-Kloosterman,4,5 Caroline TM van Rossum,5 and Marga C Ocké5

Abstract

Background: For the evaluation of both the adequacy of intakes and the risk of excessive intakes of micronutrients, all potential sources should be included. In addition to micronutrients naturally present in foods, micronutrients can also be derived from fortified foods and dietary supplements. In the estimation of the habitual intake, this may cause specific challenges such as multimodal distributions and heterogeneous variances between the sources.

Objective: We present the Statistical Program to Assess Dietary Exposure (SPADE) that was developed to cope with these challenges in one single program.

Method: Similar to other methods, SPADE can model habitual intake of daily and episodically consumed dietary components. In addition, SPADE has the option to model habitual intake from dietary supplements. Moreover, SPADE offers models to estimate habitual intake distributions from different sources (e.g., foods and dietary supplements) separately and adds these habitual intakes to get the overall habitual intake distribution. The habitual intake distribution is modeled as a function of age, and this distribution can directly be compared with cutoff values to estimate the proportion above or below. Uncertainty in the habitual intake distribution and in the proportion below or above a cutoff value is quantified with ready-for-use bootstrap and provides 95% CIs.

Results: SPADE is implemented in R and is freely available as an R package called SPADE.RIVM. The various features of SPADE are illustrated by the estimation of the habitual intake distribution of folate and folic acid for women by using data from the Dutch National Food Consumption Survey 2007–2010. The results correspond well with the results of existing programs.

Conclusion: SPADE offers new features to existing programs to estimate the habitual intake distribution because it can handle many different types of modeling with the first-shrink-then-add approach. J Nutr 2014;144:2083–91.

Keywords: SPADE, habitual intake, usual intake, shrink-then-add, micronutrients, multimodality

Introduction

In the evaluation of dietary intake of populations, one is often interested in the habitual (usual) intake, that is, the long-term average intake. In food consumption surveys, dietary intake is generally collected with short-term measurements, for example, 24-h recalls. Twenty-four-hour dietary recalls repeated on non-consecutive days can be used to estimate the habitual intake by statistical correction for within-person variation. The general idea of this correction is that first a transformation is applied to the observed data to make the distribution reflect a normal distribution. At the transformed scale, the within-person variability is removed, resulting in a shrunken distribution. Finally, a complex back transformation is applied to the original scale to make a meaningful presentation of the results (1–4).

In the past decades, several applications have become available to estimate the habitual intake distribution from repeated short-term intake measurements. These programs were developed for different domains of application, for example, food safety (5–7), nutritional monitoring (2, 8–10), and epidemiology (11). Generally, these models can be used to estimate the habitual intake distribution for dietary components consumed daily by almost all subjects in the study population and/or episodically consumed dietary components.

For the evaluation of the adequacy and risk of excessive intakes of micronutrients, all potential sources should be included.
In addition to the intake of micronutrients naturally present in foods, fortified foods and dietary supplements can also contribute substantially to the total micronutrient intake, at least for part of the population (12–14). The estimation of the habitual intake distribution of micronutrients from different sources poses some specific challenges (15). One challenge is a multimodal intake distribution that may be caused by a part of the population consuming the micronutrient in relatively large amounts from fortified foods and/or dietary supplements (16). For multimodal distributions, the transformation to a normal (symmetrical) distribution may fail with the generally applied transformation methods (16). A second challenge is heterogeneous between-person variation for micronutrient intake from the different sources (17). Consequently, the assumption of homogenous variances is violated, and this influences the habitual intake distribution estimated by adding these sources before correction for within-person variation. A third challenge is the spiked distribution of observed intake from dietary supplements (18). For this type of distributions, no transformation is available to obtain normally distributed supplement intakes.

A solution to deal with the first 2 challenges is to model the habitual intake distribution for each source of the dietary component separately. Subsequently, these habitual intake distributions are combined to obtain the total habitual intake distribution of the micronutrient (17, 19–22). This solution is called the first-shrink-then-add approach. A similar approach was also proposed to estimate the exposure of a contaminant present in a few food products (16). A regression tree model that can handle spiked distributions can solve the last challenge for the supplement intakes. As far as we know, no single program is available with all required modeling options to estimate the habitual intake distribution of micronutrients according to the proposed ideas. Therefore, the Statistical Program to Assess Dietary Exposure (SPADE)⁶ was developed.

In the actual study, SPADE is introduced, and the various modeling options in SPADE are described by estimating the habitual intake distribution of folate and folic acid with the use of data from the Dutch National Food Consumption Survey (DNFCS) 2007–2010 (18). The aim of this example was to demonstrate the 4 main options of SPADE, together with the available bootstrap functions and output-like diagnostic plots (Supplemental Figures 1–7, Supplemental Tables 1–5). SPADE results were also compared with the results of the SAS macros of the National Cancer Institute (9, 11) (Supplemental Comparison).

Like other available models, SPADE can be applied to intake of other micronutrients, other dietary components, and foods.

The SPADE Program

SPADE is a program in R (23) and estimates the habitual intake of dietary components (e.g., micronutrients, macronutrients, contaminants) or food (groups) from repeated short-term dietary intake data. SPADE consists of several modeling options. The first option models dietary components that are consumed on a daily basis by almost all participants (1-part model). A second option concerns dietary components that are consumed episodically (2-part model). The third option specifically aims to estimate total habitual nutrient intake from foods and dietary supplements (3-part model). With the fourth option, the habitual intake from multiple food sources or food groups can be estimated in a multipart model. In the third and fourth model options the first-shrink-then-add approach is used (20, 21).

Before the description of the models, some general principles of SPADE are given.

- SPADE requires a dataset with observations on 2 independent days for each subject, which means at least 2 nonconsecutive study days.
- By default, intake is modeled as a function of age; however, age-independent modeling is possible.
- Results can be presented per age unit (e.g., year, month) or in self-defined age groups.
- Normalized survey weights (including sampling strata) can be incorporated to remove bias from a survey sample and to make the results better representative for the target population.
- Direct comparison of the habitual intake distribution with cutoff values (e.g., dietary reference values) is implemented. Three types of comparisons are possible as follows: 1) proportion below, 2) proportion above, and 3) comparison of median habitual intake with a specific level. Cutoff values may vary over age units.
- Quantification of uncertainty in data and model is provided by bootstrap confidence intervals (default 95%) of all percentiles, the mean of the habitual intake distribution, and the proportions below/above cutoff values.

SPADE can be downloaded from http://www.spade.nl as the R-library SPADE.RIVM. A detailed user's manual is also available from that website.

1-Part model for daily-consumed foods or food components. The 1-part model is intended to estimate the habitual intake distribution of food components, foods, or food groups that are consumed daily by (almost) all subjects. This means that in the study population (almost) all subjects have a non-zero intake on 2 observed days. The SPADE 1-part model is an improvement of the Age-Dependent Model (2) and consists of the following steps (Supplemental Figure 1).

1.1 Transform (Box-Cox) the observed intakes to obtain a normal distribution on the transformed scale.
1.2 Model the transformed intake as a function of age; estimate all model parameters, including the within- and between-person variances; and use these to obtain the shrunken distribution on the transformed scale.
1.3 Back-transform the distribution percentiles (per age unit) from the transformed scale to the original scale and eliminate the within-person variance [modified Gaussian-Hermite Quadrature (3)].

2-Part model for episodically consumed foods or food components. The 2-part model is appropriate to estimate the habitual intake distribution of food components, foods, or food groups that are consumed episodically. The observations can typically be divided into 3 groups: subjects with positive intakes on both days, subjects with only one positive intake, and subjects with zero intakes on both days. By default, all subjects are assumed potential consumers. First, all observations are used for the estimation of the distribution of the probability of consumption. Next, for the estimation of the distribution of habitual amounts on consumption days, only subjects with at least 1 day with a positive intake are involved, because 2 zero

---

⁶ Abbreviations used: AI, Adequate Intake; DNFCS, Dutch National Food Consumption Survey; EAR, estimated average requirement, 50% of the population's requirement is met; SPADE, Statistical Program to Assess Dietary Exposure; UL, Tolerable Upper Intake Level.
intakes for a subject do not provide any information about the habitual intake amount of that subject.

The modeling step for the first part of the model (probabilities of consumption) is as follows.
1.1 Estimate the probability of consumption as a function of age with a Beta-Binomial model (24).

The modeling steps for the second part of the model (amounts on consumption days) are as follows.
2.1 Select the positive intakes of all subjects with one positive intake and select randomly one positive intake from the subjects with 2 positive intakes. Therefore, subjects with 2 zero intakes are omitted. Transform (Box-Cox) the selected intakes to obtain a normal distribution.
2.2 Model the transformed amounts as a function of age and estimate all model parameters, including the total residual variance.
2.3 Estimate within- and between-person variances on the basis of the data of subjects with 2 non-zero intakes and the fitted model of 2.2.
2.4 Apply the ratio of both variances obtained in 2.3 to the total residual variance of 2.2 as an estimate for the within- and between-person variances in 2.2.

With these 2 variances the model fitted in 2.2 provides all information for the back-transformation.

The habitual intakes are derived as the probability on consumption (part 1) multiplied by the habitual intake amount on a consumption day (part 2), which are obtained by a Monte Carlo simulation.

Never-users (or never-consumers) are persons who indicated that they never consume a specific food or dietary supplement. When this information is available in the data, the previous 2-part model in SPADE considers this.

Potential users (or potential consumers) are all other users who are not never-users. Potential users may have zero intakes on all survey days, but they still have a positive probability of having a positive intake.

2-Part model for intake from dietary supplements. The distribution of observed intake from dietary supplements may be spiked. In this case, the previous described 2-part model cannot be used. Therefore, an alternative 2-part model is needed. First, if information on never-users of dietary supplements is made available, SPADE selects the never-users and assigns a habitual intake of zero to them. Next, SPADE continues with the potential users of dietary supplements with the compound of interest and estimates the probability of taking the supplement as in the 2-part model for food and food components. In the DNFCs, potential users of dietary supplements were defined as follows: subjects who reported intake of a (specific nutrient from) supplement in at least one of the 24-h recalls and/or subjects who reported (specific nutrient) supplement use in the frequency questionnaire. As a result, for some subjects no amount information was available (user according to frequency questionnaire, but no intake during 24-h recalls).

The second step is as follows. For subjects with 2 observed non-zero intakes from dietary supplements, the habitual intake amount is set as the mean intake during both days, because for most subjects with 2 non-zero intakes, these intakes are the same (17). For subjects with only one observed positive intake from dietary supplements, this single positive value is used as the habitual intake amount. For the remaining subjects, the potential dietary supplement users with 2 zero intakes on both survey days, a habitual intake amount is predicted with a regression tree model that is based on the data of all subjects with a habitual intake amount. The regression tree model [explained in Breiman et al. (25) and Therneau and Atkinson (26)] determines subpopulations on the basis of covariates (e.g., age, type of supplement used, frequency of supplement use), minimizing the sum of subpopulation variances. Next, this regression tree is used to predict for each potential dietary supplement user with 2 zero observations in the user’s subpopulation. The habitual intake amount of this user is set to a randomly chosen habitual intake amount of this subpopulation.

The habitual intake distribution for a representative sample of the general population needs to be constructed from the original study population sample. This distribution is obtained by generating 100 pseudo persons for each person in the survey. For the never supplement users, the pseudo persons have intakes of zero. For the ever supplement users, the 100 pseudo persons get a habitual intake amount randomly drawn from the subset in which they belong according to the regression tree. In addition, each pseudo person randomly receives an average intake probability, according to the fitted Beta-Binomial model. In this way, the means of the habitual intake amounts and probabilities of the 100 pseudo persons reflect the corresponding means of the original person, and the between-person variation is taken into account. These habitual intakes of a compound from supplements for pseudo supplement users are joined with the zeroes of pseudo never-users to obtain the population habitual intake distribution of the compound from supplements.

3-Part model for total intake from food and dietary supplements. Habitual intake from food and dietary supplements can be estimated by simply summing the amounts before correcting for within-person variation with subsequent application of a 1- or 2-part model, a so-called first-add-then-shrink approach. This approach does not take into account potential differences in model parameters or variances between the intake from food and dietary supplements. In addition, such a combined distribution may have a multimodal shape. To overcome these issues, a 3-part model for the estimation of habitual total intake distributions from food and dietary supplements was proposed by Verkaik-Kloosterman et al. (17) (Figure 1). The general idea of this method is to estimate first the habitual intake from food only separately in never supplement users (part 1) and in supplement users (part 2) and to estimate the habitual intake from dietary supplements (part 3). Next, pseudo persons are generated for the 3 parts. Thereafter for the supplement users, the habitual intakes from food and from supplement are added per pseudo person (the add step) to obtain the total habitual intake. Finally, the complete habitual intake distribution is obtained by joining the habitual intakes from the pseudo never-supplement users with habitual intakes from the pseudo supplement users. This first-shrink-then-add approach is illustrated in Figure 1.

Multipart model for total intake from 2 or more food sources. The first-shrink-then-add approach may also be required to estimate the habitual intake distribution from several food sources. An example is the estimation of habitual intake of a specific compound (e.g., a contaminant) present in a few specific episodically consumed foods for which a first-add-then-shrink approach would result in a multimodal distribution [explained in Slob et al. (16) and Kennedy (22)]. Another example is a nutrient naturally present in foods in small amounts and present in fortified foods in larger amounts. For all different sources (supplements, one or more food sources, or one food source with known never-users), the habitual intake distributions

Statistical program to assess dietary exposure 2085
are estimated separately with the use of the appropriate models (Figure 2). For practical reasons, in the multipart model the habitual intake of the sources other than dietary supplements is not estimated separately for supplement users and nonusers. The habitual total intake distribution is obtained by sampling pseudo persons for each subject in the survey.

**Data used for the examples.** The food consumption data in the examples are from female participants in the DNFCS 2007–2010 (18). Briefly, this survey was conducted among persons 7–69 y old (n = 3819; female participants, n = 1908) who were drawn from representative consumer panels. The study was conducted according to the guidelines of the Helsinki Declaration of 1975, as revised in 1983, 1989, 1996, and 2000. Data were collected with a general questionnaire and two 24-h dietary recalls on independent days (2- to 6-wk interval). The questionnaire included questions on the consumption frequency of dietary supplements. For the presented examples, users of folic acid–containing dietary supplements were defined as subjects that indicated use of folic acid supplements, multivitamins, or multimineral/multivitamins. For the 24-h recalls of foods, beverages, and dietary supplements consumed, the computer-directed interview program EPIC-Soft was used (27, 28). To estimate nutrient intake, the consumption data were combined with data from the extended version of the Dutch Food Composition Database [NEVO table 2011/3.0 (29)] and data from the Dutch Supplement Database, dated 1 January 2008 (30). The total folate intake was expressed in folate equivalents, whereby 1 μg of folate from food = 1 μg folate equivalent, 1 μg of folic acid from food = 1.7 μg folate equivalent, and 1 μg of folic acid from dietary supplements = 2 μg folate equivalent (31).

To illustrate the assessment of folate intake adequacy, the estimated average requirement (EAR) and the Adequate Intake (AI) for folate equivalent intake from the Health Council of the Netherlands were used (31). The percentage of the population with inadequate intake was assessed with the use of the EAR cut-point method. For age groups for which no EAR was available, a qualitative assessment was made with an AI (32). The Tolerable Upper Intake Level (UL) for folic acid, as set by Scientific Committee on Food (33), was applied to estimate the proportion of the population potentially at risk of excessive intake of folic acid.

To get results representative for the Dutch population all analyses were weighted for small deviations in sociodemographic characteristics (level of education, region, and urbanization), day of the week, and season of data collection first recall. In addition, results across ages were weighted for age with the use of Dutch population numbers 1 January 2008 (34).

All analyses were performed with R version 3.1.0 and SPADE.RIVM version 2.32.06 (R-library). Four models of SPADE and several general features and modeling options are illustrated with the use of relevant examples of folate or folic acid intake from different sources. An overview is presented (Table 1).

**Results**

**Example 1, the 1-part model: folate intake from food.** All female subjects in DNFCS 2007–2010 consumed folate daily. The observed folate intake distribution was skewed to the right (Supplemental Figure 2, left side), with a 5–95th percentile range of 80–361 μg/d and a mean of 196 μg/d. After a Box-Cox transformation with λ = 0.27 (95% CI: 0.23, 0.31), the distribution was more or less normal (QQ-plot shows a straight line), with some deviations at both tails (Supplemental Figure 2, right side). In addition, the distribution of the residuals (i.e., the difference between the transformed observed intakes and the predicted intakes on the transformed scale by the model) was similar to the normal distribution with some deviations at both tails (Figure 3). On the λ = 0.27 scale, the estimates of the intercept, coefficient for (age/10)^0.5, and between- and within-person variances were 13.01, −2.44, 0.9, and 2.3, respectively.

For each year of age, the habitual intake distribution was estimated, and uncertainties were quantified (Figure 4, Supplemental Table 1). The uncertainty was largest at the right tail of the distributions, observed as the widest 95% CI. Especially

![FIGURE 1](https://example.com/figure1.png)

**FIGURE 1** Scheme of the SPADE 3-part model. Never-users and users refer to the never-users and the users of the supplements. The models for the 2 food parts (never-users and users) may be different. The add step applies only to pseudo persons in the supplement users part. SPADE, Statistical Program to Assess Dietary Exposure.

![FIGURE 2](https://example.com/figure2.png)

**FIGURE 2** Scheme of the multipart options in SPADE with first shrink (the separate modeling in the upper parts) then add (for each pseudo person in the lower part) approach. For the food with never-users and for the supplement part, external information (e.g., from a FFQ), is needed. hab., habitual; SPADE, Statistical Program to Assess Dietary Exposure.
among the younger ages, folate intake increased by age. The median intake ranged from 141 μg/d (95% CI: 136, 156 μg/d) for girls 7 y old to 229 μg/d (95% CI: 221, 241 μg/d) for women 68 and 69 y old.

**Example 2, the 2-part model: folic acid intake from food.**

One source of folic acid is fortified foods. Currently in The Netherlands, foods may be voluntarily fortified with folic acid to a maximum of 100 μg/100 kcal (35). Folic acid fortification is not common in The Netherlands (36); consequently, the intake is episodical. In our study population, 1294 female subjects (68%) had no folic acid intake on both recalled days, 259 women (14%) consumed folic acid from food sources on both recalled days, and 355 women (19%) had an intake on one study day. The histogram of the observed mean intake per person (in μg/d) shows the 68% with 2 zero intakes is visible in the most left bin (Supplemental Figure 3).

In the first part of the model, the intake probabilities were estimated as a function of age. A diagnostic plot (Supplemental Figure 4) is given to check some properties of the residuals. The 2 upper plots of data in Supplemental Figure 4 help to check unexplained relations between the residuals and fitted values (left side) or between residuals and age (right side). In the example, the residuals show a random behavior between the values –4 and 4 for both the fitted (left) and age (right) values, suggesting no unexplained relations. The density and QQ-plot of the quantile residuals corresponded with a normal distribution, with some deviations at both tails (Supplemental Figure 4).

In the second part of the model, the habitual intake amounts given consumption were estimated. On the λ = 0.21 scale, the estimates of the intercept, coefficient for age, and between- and within-person variances were 6.22, 0.009, 1.5, and 1.9, respectively. The estimated probability to consume folic acid from fortified foods (part 1; Supplemental Figure 5A) and the estimated habitual intake amount of folic acid given consumption (part 2; Supplemental Figure 5B) were combined to get an estimation of the habitual intake distribution of folic acid (Supplemental Figure 6, Supplemental Table 2). The median habitual intake of folic acid ranged from 4 to 15 μg/d and increased with age. The lower percentiles were close to zero (Supplemental Table 2), as expected for an episodically consumed nutrient with the observed data of 68% with no intake during the survey days. However, the median of the habitual intake distribution per age class was positive (Supplemental Table 2). Because no information on never-consumers was available, all participants were supposed to be potential consumers and had a positive probability (Supplemental Figure 5A) on having a positive habitual intake amount.

**Example 3, the 3-part model: folic acid intake from food and dietary supplements.**

Besides fortified foods, dietary supplements are also a source of folic acid. To evaluate if folic acid intake was too high, habitual folic acid intake from both sources should be compared with the UL. As explained in the general explanation of SPADE (3-part model), a first-shrink-then-add approach was most appropriate to estimate the total intake from food and supplements. With this approach, potential differences in within- and between-person variances and model parameters between intake from foods in supplement users, intake from foods in non-supplement users, and intake from dietary supplements are accounted for. Consequently, potentially multimodality caused by the intake from different sources could be retained.

For never supplement users (1147 women 7–69 y, 60% of all women in this age class) the observed intake data from food sources were Box-Cox transformed with λ = 0.29 (95% CI: 0.19, 0.39) and the within- and between-person variances were, respectively, 3.62 and 2.37. For potential users of folic acid–containing dietary supplements (761 women 7–69 y, 40% of all women in this age class), the observed intake data from food sources were transformed with λ = 0.26 (95% CI: 0.15, 0.38), and the estimated within- and between-person variances were, respectively, 2.71 and 2.58.

For 349 female subjects, mean folic acid intake from dietary supplements in the 24-h recalls was greater than zero. Supplemental Figure 7 shows a histogram of all positive supplement intakes, which shows a spiky density. Whereas 412 female subjects had indicated they took folic acid–containing dietary supplements in the general questionnaire, they reported no dietary supplements that contained folic acids on both recalled days. For these latter female subjects, a regression tree model was used to predict their habitual dietary supplements amount.

![FIGURE 3 Screen shot of a diagnostic SPADE plot to check for normality of residuals of the fitted model for folate intake, female subjects, 7–69 y (Dutch National Food Consumption Survey 2007–2010). SPADE, Statistical Program to Assess Dietary Exposure.](https://academic.oup.com/jn/article-abstract/144/12/2083/4644413)
We therefore used the 2-part model for supplements, including the regression tree modeling with the intake frequency of folate-containing supplements in winter and in summer from the food-frequency questionnaire as explanatory variables. The regression tree divided the 349 female subjects into 5 subgroups (the end nodes of the regression tree) and assigned them positive habitual folic acid supplement intake means with intakes in the range of 189–450 μg/d (Figure 5). Subsequently, for subjects with 2 zero folic acid intakes from dietary supplements in both 24-h recalls, a habitual intake was predicted by taking randomly a habitual supplement intake from their predicted subgroup (Figure 5).

The median total habitual folic acid intake from both fortified foods and dietary supplements was highest for women aged 51–69 y, namely 28 μg/d (95% CI: 21, 38 μg/d), whereas for girls 7–18 y the median habitual intake ranged from 7 to 10 μg/d (Supplemental Table 3). Comparison with the age-specific UL showed that <0.5% of the female population had intakes above the age-specific UL.

Further nutritional results on dietary supplements can be found in the DNFCS 2007–2010 (18).

**Example 4, the multipart model for folate equivalent intake from food, fortified foods, and dietary supplements.**

To estimate the adequacy of folate intake, the folate and folic acid intakes from all sources were combined and expressed as folate equivalents. The sources were 1) folate from foods, 2) folic acid from fortified foods, and 3) folic acid from dietary supplements. Supplemental Table 4 shows some characteristics of the density of the observed folate equivalent intake from the 3 sources.

The results of example 1 and 2 showed that the between- and within-person variations differed for folate naturally present in food (the between- and within-person variances were 0.9 and 2.3, respectively) and for folic acid present in fortified foods (1.5 and 1.9, respectively). Moreover, by adding the intake from foods to the intake from fortified foods, the intake distribution changed to a distribution with a larger tail to the right. This influenced the transformation parameters (data not shown). Therefore, a first-shrink-then-add approach was applied to obtain total habitual folate equivalent intake. The median habitual total folate equivalent intake increased with age from 174 μg/d (95% CI: 169, 189 μg/d) for girls 7–8 y to 302 μg/d (95% CI: 287, 321 μg/d) for women 51–69 y (Supplemental Table 5). Approximately 10–27% of the adult women had a total habitual folate equivalent intake below the EAR, indicating that up to one-fourth of the women had inadequate folate equivalent intake. For girls comparisons were made with the AI, because no EAR was derived by the Health Council of The Netherlands. For girls aged 7–8 y the risk of inadequate intake was low (i.e., median > AI), whereas for older girls no statement about adequacy can be made, because the median intake was lower than the AI.

An overview is presented of folate equivalent intake (in μg/d) for female subjects 7–69 y for the 3 different folate sources and for the total folate equivalent habitual intake estimated by the first-add-then-shrink method and by the first-shrink-then-add method, respectively (Table 2). The distribution of the habitual intake estimated with the first-add-then-shrink method is not equal to the distribution estimated with the first-shrink-then-add method. For some percentiles, the estimated habitual intake was significantly different between both methods (e.g., Table 2, 5th, 50th, 75th, and 95th percentiles). However, the estimates at other percentiles did not differ significantly (e.g., Table 2, 25th percentile) between both methods.

**Discussion**

In this study, we introduced SPADE to obtain habitual intake distributions from short-term intake measurements. We found that, next to the basic modeling of habitual intake, SPADE was used for the estimation of the between- and within-person variations. The effects of the transformations were therefore determined by the between- and within-person variances.

**FIGURE 5** Screen shot of SPADE. Fitted regression tree is based on the habitual folic acid intake from dietary supplements for users (female subjects 7–69 y) with at least one positive observation and with external information of an additional questionnaire as covariables. In each node, mean folic acid intake from supplements and the number of women (n) in the subgroup are given (Dutch National Food Consumption Survey 2007–2010). fol_rest, intake frequency of dietary supplement folic acid outside the winter; fol_win, intake frequency of dietary supplement folic acid in winter; mult_res, intake frequency of dietary supplement (multivitamins without minerals) outside the winter; SPADE, Statistical Program to Assess Dietary Exposure; vitb_res, intake frequency of dietary supplement vitamin B complex outside the winter.
specifically designed to address challenges in modeling habitual intake for micronutrients. These challenges include spiked distributions of micronutrient dosages in dietary supplements, multimodal intake distributions, and heterogeneity in variances because of different sources of micronutrients (natural foods, fortified foods, and dietary supplements). A good estimation of total micronutrient intake is relevant because dietary supplements and fortified foods are substantial sources of various micronutrients (18, 37). For most micronutrients, the total habitual intake distribution of a population should be compared with dietary reference values to estimate nutrient adequacy and the potential for excessive intakes in a population (15).

In SPADE, the challenges of multimodal intake distributions and heterogeneity in variances were dealt with by modeling habitual intake separately for different sources and adding them afterward rather than the other way around. The chosen approach (20, 21) is also called the first-shrink-then-add approach and was applied by Verkaik-Kloosterman et al. (17) and others (16, 22, 38). However, to our knowledge no habitual intake estimation program is currently generally available which allows application of this approach for intakes from foods and dietary supplements. In addition, specific to SPADE is the modeling of intake from dietary supplements with a spiky density with the use of a regression tree model. As far as we know, no other usual intake software is available with this approach implemented.

To illustrate the different model options, we applied SPADE to intake data of folate and folic acid from different sources. In addition, others estimated usual intake of folate equivalents naturally available in foods, added to foods, and from dietary supplements. Some researchers modeled usual intakes of folate equivalents naturally present and added to foods with a first-add-then-shrink approach (13, 14, 37), whereas others used the first-shrink-then-add approach (37–40). To our knowledge no other researchers used the first-add-then-shrink approach to add folate equivalents from dietary supplements to this intake. This can however be explained because dietary supplement consumption was assessed for the previous month (13, 14, 37–40) rather than for 2 specific days as in our food consumption survey.

Several other programs and models exist to estimate habitual intake, comparable with the 1-part and 2-part models of SPADE. Souverein et al. (41) compared 4 programs [Iowa State University/Software for Intake Distribution Estimation (42), National Cancer Institute (9), Multiple Source Method (10), and an earlier version of SPADE] to estimate habitual intake of daily and episodically consumed nutrients and foods. Because the methods provided comparable results for the data in this study, they concluded that the preference for one method above the other might be due to different features in the software and aspects such as availability, costs, and user friendliness or other practical reasons.

Positive features of SPADE are the options to model intake as a function of age per unit and per age group, the option to obtain 95% CIs, the possibility to include survey weights, and the possibility to compare intake distributions with the EAR, AI, or UL according to the methods described by the US Institute of Medicine (43). Second, user friendliness was an important criterion in the development of SPADE. User friendliness included easy handling of estimations of habitual intake distributions for multiple dietary components at once, clear and easy-to-read diagnostic plots, flexibility in input and output formats, and a detailed manual. Third, SPADE was developed in R (23), which is freely available. R was chosen to create easy accessibility of SPADE and to stimulate collaboration with other researchers for further development.

Apart from the above strengths of SPADE, limitations should also be acknowledged.

SPADE was developed for nutritional applications to the DNFCS. For this reason, SPADE may not be directly applicable to surveys with different study designs (e.g., to studies with ≥3 or a varying number of observation days among study participants) or to studies in which cluster sampling was used. We foresee future adjustments to SPADE to make it more generally applicable to other surveys.

Unlike some other models (7, 9, 11, 22, 44, 45) SPADE does not allow one to take correlations between intake probability and intake amount into account. However, simulation study (46) showed that for a low correlation or for large intake frequencies, it is not advantageous to include a correlation into the model. In other situations, the study showed less bias when the correlation was taken into account. Therefore, further development of SPADE in this area is warranted. Such development for the 3-part and multipart models will probably need Bayesian solutions that will enable the user to account for all the possible correlations [explained in Kennedy (22) and Zhang et al. (45)]. A general limitation of all models to estimate the habitual intake of food components is the potential bias in the underlying dietary data (1). The models cannot remedy this bias.

In conclusion, SPADE offers some new features to existing programs to estimate the habitual intake distribution because it can handle many different types of modeling. In addition to the

---

**TABLE 2** Distributions of habitual folate equivalent intake from different sources (foods, fortified foods, dietary supplements, and all sources) with the shrink-then-add and the add-then-shrink approaches for female subjects 7–69 y (DNFCS 2007–2010)³

<table>
<thead>
<tr>
<th>Source, µg/d</th>
<th>Mean intake</th>
<th>5th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>95th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>210</td>
<td>131</td>
<td>173</td>
<td>206</td>
<td>242</td>
<td>301</td>
</tr>
<tr>
<td>Fortified foods</td>
<td>32</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>47</td>
<td>126</td>
</tr>
<tr>
<td>Dietary supplements</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>364</td>
</tr>
<tr>
<td>Method, µg/d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total shrink-then-add</td>
<td>319 (309, 336)</td>
<td>149 (145, 154)</td>
<td>207 (203, 213)</td>
<td>264 (257, 273)</td>
<td>360 (346, 381)</td>
<td>668 (635, 730)</td>
</tr>
<tr>
<td>Total add-then-shrink</td>
<td>315 (305, 327)</td>
<td>133 (128, 139)</td>
<td>211 (205, 217)</td>
<td>287 (278, 296)</td>
<td>388 (374, 404)</td>
<td>593 (562, 631)</td>
</tr>
</tbody>
</table>

³ DNFC, Dutch National Food Consumption Survey.

Values are the mean and 5 percentiles of the habitual intake distribution estimated by the indicated method.

Dietary supplements 60 0 0 0 30 364

Food 210 131 173 206 242 301

Fortified foods 32 0 1 12 47 126

Dietary supplements 60 0 0 0 30 364

Method, µg/d

Total shrink-then-add 319 (309, 336) 149 (145, 154) 207 (203, 213) 264 (257, 273) 360 (346, 381) 668 (635, 730)

Total add-then-shrink 315 (305, 327) 133 (128, 139) 211 (205, 217) 287 (278, 296) 388 (374, 404) 593 (562, 631)

---

In conclusion, SPADE offers some new features to existing programs to estimate the habitual intake distribution because it can handle many different types of modeling. In addition to the
basic modeling of daily and episodically consumption, SPADE applies a first-shrink-then-add approach to combine the intake from different sources to overcome problems with multi-modality and heterogeneous variances. Although SPADE was developed to estimate the habitual intake of micronutrients, SPADE is also applicable to other food components, foods, and food groups.

Acknowledgments

ALMD and JV-K designed the research; ALMD developed SPADE; JV-K analyzed the data; CTMvR provided the data; ALMD and JV-K designed the research; ALMD developed

References

7. Boer W, van der Voet H. MCRA 7 a web-based program for Mont Carlo Risk Assessment. Wageningen and Bihoven (The Netherlands); Biomixis (Wageningen University and Research Centre) and RIVM (Dutch National Institute Public Health and the Environment); 2011.
20. Garriguet D. Challenges in combining food and supplement intake data using the Canadian Community Health Survey - nutrition. 7th International Conference on Diet and Activity Methods (ICDAM 7); 2009 June 4–7; Washington, DC. Washington (DC): ICDAM; 2009: 132.