J. Nutr. 136:2916-2920, November 2006

The Potential of AGE MODE, an Age-Dependent Model, to Estimate Usual Intakes and Prevalences of Inadequate Intakes in a Population[1]

Patricia M. C. M. Waijers[2,*], Arnold L. M. Dekkers[3], Jolanda M. A. Boer[2], Hendriek C. Boshuizen[3] and Caroline T. M. van Rossum[2]

[2] Center for Nutrition and Health and [3] Expertise Center for Methodology and Information Services, National Institute for Public Health and the Environment, Bilthoven, the Netherlands

[*] To whom correspondence should be addressed. E-mail: patricia.waijers@rivm.nl

ABSTRACT

Dietary intake data often stem from short-term measurements. However, for dietary assessment, generally the habitual intake distribution is of interest. Currently, habitual intake distributions are often estimated separately for subgroups of gender and age and do not take into account the variation in intake caused by age within age groups. Therefore, we developed an age-dependent dietary assessment model, which was demonstrated and tested using folate intakes from the third Dutch National Food Consumption Survey, conducted in 1997/98. The proposed model produced estimates of the mean habitual intake and intake percentiles as a function of age. The methodology has clear advantages in estimating habitual intakes in children. Also, given the large variation in intakes of several dietary components, estimated habitual intakes produced by other methods may have low precision and be less reliable if numbers are small. In our age-dependent model, all available data can be used to estimate the parameters of the habitual intake distribution, improving the precision of the estimates, and providing consistent estimates for a larger population sample as no subgroups need to be created. Although the model may still be further developed, the feature of age dependency shows clear advantages above methods currently used to estimate habitual intakes.

Introduction

Food consumption surveys are conducted to gain insight into the food consumption of, and nutrient supply to, a population. A 24-h recall on at least 2 nonconsecutive days was concluded to be the most suitable method to obtain internationally comparable data on population mean actual intakes (1). But, in contrast to food frequency data, for example, dietary recalls provide actual or observed dietary intake data, which include day-to-day variations (2). For dietary assessment at the population level, nutritionists are generally interested in the habitual intake distribution, reflecting individuals' long-term mean intake of dietary components in a population (3). Statistical methods have been proposed for estimating and eliminating the intra-individual variation from short-term measurements for subgroups of gender and age (4–7). However, by specifying subgroups, obvious variations in intake may be ambiguous, as in children the variation in intakes can to a large extent be explained through age. Subgroups are specified by the researcher and resulting group intake estimates are averages of children of various ages, whereas a relation with age is expected. In addition, numbers of individuals in the distinct subgroups may be small, affecting the reliability of the estimated habitual intake distribution, especially in the presence of high variations in intakes.

Therefore, we propose an age-dependent model to estimate habitual intakes. In an age-dependent model, habitual intake distributions in a population are estimated and described as a function of age. Also, consecutive steps in dietary assessment can be performed age-dependently, for example, relating habitual intakes to estimated average requirements (EARs)[4] or upper safe levels of intake by estimating the proportion of individuals for whom intake is inadequate or too high.

We developed an age-dependent dietary assessment model, called AGE MODE, which is evaluated and compared with currently used methods. We showed the usefulness of this method using folate intakes from the third Dutch National Food Consumption Survey (DNFCS-3), conducted in 1997/98.

Methods

Description of the model

AGE MODE is a general methodology that can be used for intakes of micronutrients and other dietary compounds. It aims to estimate habitual intakes of dietary components in a population from short-term dietary data. In contrast to other proposed methodologies, this is accomplished by describing intakes as a continuous function of age. The methodology is based on ideas of Slob (7) and generalized as much as possible. The model has been programmed in S-PLUS (8).

AGE MODE contains several steps that will be explained in more detail later in this paper: 1) Box-Cox transformation of the observed intakes to obtain normally distributed data; *2*) fitting a fractional polynomial to the transformed data; *3*) obtaining a mixed effect estimate of the fractional polynomial, providing the inter-individual variance and the intra-individual variance; *4*) identification of possible outliers; *5*) back-transformation by Monte Carlo Simulations to obtain the habitual intakes on the original scale; and *6*) additional steps in dietary assessment.

Box-Cox transformation of the observed intakes to obtain normally distributed data. Intake data are generally skewed, whereas most statistical analyses require normally distributed data. The Box-Cox method estimates the transformation parameter λ using the maximum likelihood method, ensuring symmetrically and approximately normally distributed intake data after transformation (Eq. 1).

$$f(x)=(x^{\lambda}-1)/\lambda. \qquad \text{(Eq. 1)}$$

Note that for $\lambda = 0$, f(x) = ln(x).

Fitting a fractional polynomial to the transformed data. AGE MODE searches the best polynomial function to describe the data using fractional polynomial regression of order 2 (9).

Let *n* denote the number of observations and let *p* and *q* be the powers of the fractional polynomial $y(x_i)$. The fractional polynomial regression function (Eq. 2) is given by

$$y_i=a+b\cdot(x_i)^p+c\cdot(x_i)^q+\varepsilon_I\ (i=1,2,\ldots,n,\ p\neq q)$$
$$\text{or}$$
$$y_i=a+b\cdot(x_i)^p+c\cdot(x_i)^q\cdot\ln(x_i)+\varepsilon_I\ (i=1,2,\ldots,n,\ p=q), \quad \text{(Eq. 2)}$$

where x_i is the age of individual i, and y_i the transformed intake; p and q can take the value of {−2, −1, −0.5, 0, 0.5, 1, 2}. In this way, the transformed intakes are described as a function of age by at most a 3-parameter family of curves, and the optimal fractions p and q are estimated as well as a, b, and c.

Obtaining a mixed-effect estimate of the fractional polynomial, providing the inter-individual variance and the intra-individual variance. Because intake data for each person for at least 2 d are available, Equation 2 is refit with a mixed effect model. Each person is seen as a group with 2 or more observations, allowing estimation of the intra-individual day-to-day variance, τ^2, and the inter-individual variance, denoted by σ^2. We redefine Equation 2 for the case $p \neq q$ and define a as a random parameter, the individuals being the grouping variable. The intakes for an individual may differ by day so that Equation 2 can be reformulated into:

$$y_{ij}=a+\alpha_i+b\cdot(x_i)^p+c\cdot(x_i)^q+\varepsilon_{ij}, \quad \text{(Eq. 3)}$$

where y_{ij} is the transformed intake for individual i on day j, $\alpha_I \sim N(0, \sigma^2)$, σ^2 being the inter-individual variance, and $\varepsilon_{ij} \sim N(0, \tau^2)$, τ^2 constituting the intra-individual variance. The residuals ε_{ij} are assumed to be normally distributed with constant variance over age.

Identification of possible outliers. Outliers can seriously influence estimates. Therefore, Grubbs' method is used to automatically detect outliers, assuming that the residuals of Equation 3 are normally distributed (10,11). To check if the residuals are normally distributed, diagnostic plots of the Kolmogorov-Smirnov goodness of fit test from the S-PLUS module Environmental Stats are used (12).

Check of λ. The Box-Cox transformation is used again, because the residuals, without the outliers, should be at least symmetrically distributed. The estimated Box-Cox parameter, reported as "λ-check," should be near 1, which means that no additional transformation is needed.

Iterations of steps 1–5. Outliers can influence the Box-Cox estimate, the powers of the fractional polynomial, and the estimates of the mixed-effect model. Therefore, if outliers have been removed, steps 1–5 need to be repeated until no further outliers are detected. Two or 3 iterations seem to be sufficient in practice.

Back-transformation by Monte Carlo Simulations to obtain the habitual intakes on the original scale. To obtain the usual intake for each individual, the intra-individual variance needs to be eliminated. Monte Carlo Simulations are performed to acquire the habitual intake distributions on the original scale, using the results of the fitted mixed-effect model.

Simulated intakes are generated on the transformed scale by drawing n individuals of each age and creating a time series of k intake days for each individual *i* with:

$$\alpha \sim \mathrm{N}(0, \sigma^2)$$

$$\varepsilon_{i,t} \sim \mathrm{N}(0, \tau^2) \text{ for } t=1, 2, \ldots, k,$$

resulting in a time series for each individual *i*

$$y_{i,t} = a + b \cdot (x_i)^p + c \cdot (x_i)^q + d_i + e_{i,t}, \qquad \text{(Eq. 4)}$$

where d_i, and $e_{i,t}$ are realizations of α_I and $\varepsilon_{i,t}$ in Equation 3.

Then, each of the generated observations is back-transformed to the original scale:

$$x_{i,t} = (\lambda \cdot y_{i,t} + 1)^{1/\lambda}. \qquad \text{(Eq. 5)}$$

The habitual intake distribution of a given age, with a corresponding CI, can subsequently be estimated from the individual mean intakes, averaging over *t*.

$$x_i = \mathrm{mean}(\{\lambda \cdot y_{i,t} + 1\}^{1/\lambda}). \qquad \text{(Eq. 6)}$$

For every age, the distribution of *n* individuals is given by $\{x_1, x_2, \ldots, x_n\}$, which implies that the mean of the population and all quantiles can be calculated.

Additional steps in dietary assessment. Generally estimating habitual intake distributions is only a first step in dietary assessment. Additional steps in dietary assessment can also be accomplished in AGE MODE. Population intakes are, for example, evaluated through comparison with dietary reference intakes or upper safe levels of intake. If information on required levels of intake is provided, it is straightforward to estimate the fraction of the population with a habitual intake above or below the requirements. Most straightforward is the use of a cut-off value, but also a probabilistic approach can easily be executed in our model applying Monte Carlo Simulations if a requirement distribution has been specified.

Application of AGE MODE to folate intakes from DNFCS-3

To illustrate the model, AGE MODE has been applied to estimate habitual folate intakes from the DNFCS-3. Resulting habitual folate intake estimates are also compared with estimates obtained with the method developed at Iowa State University (ISU) by Nusser et al. (6).

The DNFCS-3 data

DNFCS-3, carried out in 1997/98, comprised 6250 noninstitutionalized persons aged 1–97 y in 2564 households selected from a stratified random sample in the Netherlands (13). Because pregnancy and lactation may affect dietary habits, pregnant and lactating women were excluded. Analyses were restricted to individuals until the age of 70 y, as the number of older individuals was very small. In total, 5744 subjects (2716 men and 3028 women) remained for analyses.

Information on food consumption was obtained with a 2-d dietary record on 2 consecutive days. The foods consumed at home were recorded in a household diary for all individual members of the household by the person usually engaged in preparation of the meals. Consumption away from home was recorded by every participant in a personal diary (children <13 were assisted by 1 or both parents). Food consumption data were collected during 40 wk/y and evenly distributed over the seasons and the 7 d of the week.

Folate intake was calculated using the 2001 Dutch food composition table (14). For 179 products that were regularly consumed, folate content was missing. For these products, folate content was estimated through comparison with similar products.

Habitual intakes by the ISU method developed by Nusser et al. (6) were calculated with the software package C-SIDE (15). For this purpose, gender and age groups generally used in the Netherlands (1–3 y, 4–8 y, 9–13 y, 14–18 y, 19–50 y, and 51–70 y) were specified. For a fair comparison of results from the ISU method and results from AGE MODE, we removed outliers identified by AGE MODE from the data.

Folate intake assessment with AGE MODE

The steps in AGE MODE, as described above, were carried out. Three iterations were sufficient, as no additional outliers were removed in iteration 3 (**Table 1**). AGE MODE displays a histogram, a cumulative density function, and a QQ-plot (Supplemental Fig. 1). The large number of observations made the Kolmogorov-Smirnov test significant, but the QQ-plot showed satisfactory results. Also, "λ-check" (Table 1) differed only slightly from 1 (no additional transformation needed). AGE MODE produced various plots, showing the fitted polynomial through the intakes on the transformed scale (Supplemental Fig. 2) and the estimated habitual folate intakes as a function of age after back-transformation (**Fig. 1**). The mean observed folate intakes of all individuals of a given age are depicted in the same figure to get clear insight into the fit of the polynomial function. The resulting habitual intake distributions become slightly wider with increasing age (**Fig. 2**).

J. Nutr. 136:2916-2920, November 2006

TABLE 1 Model estimates from AGE MODE for 3 subsequent iterations when estimating the habitual folate intake as a function of age for men (1–70) in DNFCS-3

Iteration	Model estimate
Round 1	
Observations, *n*	5432
Individuals, *n*	2716
Box-Cox transformation parameter, λ	0.1222392
Function of fractional polynomial with lme	$y \sim 7.7 - 1.4 \cdot \text{age}^{-1} - 2.6 \cdot 1/\text{age} \cdot \log(\text{age})$
Number of outliers	11
λ-check[1]	1.10
Round 2	
Observations, *n*	5410
Individuals, *n*	2705
Box-Cox transformation parameter, λ	0.1880422
Function of fractional polynomial with lme	$y \sim 6.4 + 0.8 \cdot \text{age}^{0.5} - 0.06 \cdot \text{age}$
Number of outliers	1
λ-check[1]	1.02
Round 3	
Observations, *n*	5408
Individuals, *n*	2704
Box-Cox transformation parameter, λ	0.1937614
Function of fractional polynomial with lme	$y \sim 6.5 + 0.8 \cdot \text{age}^{0.5} - 0.06 \cdot \text{age}$
Number of outliers	0
λ-check[1]	1.01

[1] This is the estimated λ for the transformed intakes. Ideally, this value should be 1.

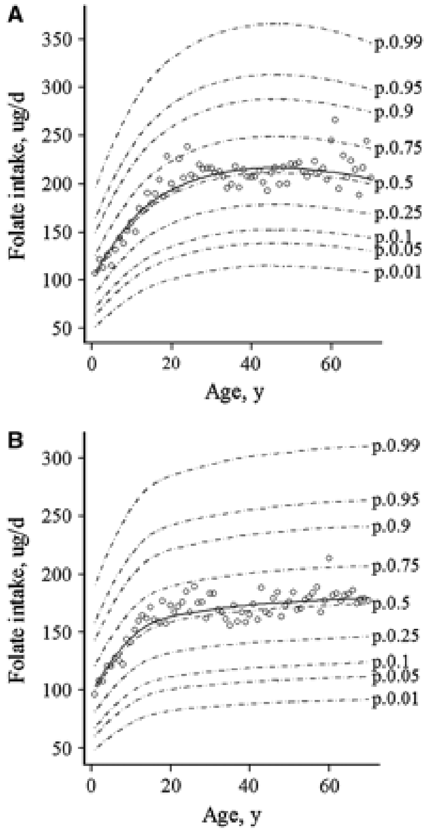

Figure 1 Habitual folate intakes as a function of age for males (*A*) and females (*B*), aged 1 to 70 y, in DNFCS-3. The solid line represents the mean of the habitual intake distribution and the dotted lines represent percentiles p1, p5, p10, p25, p50, p75, p90, p95, and p90. The dots are the mean observed folate intakes for all individuals for each age. All identified outliers were removed.

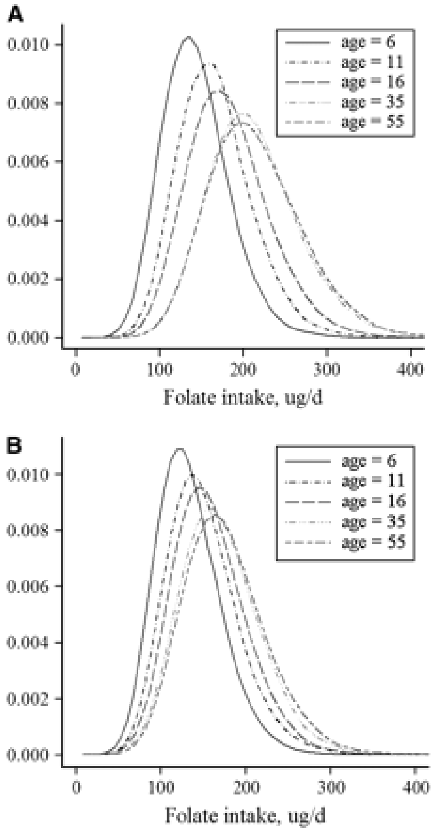

Figure 2 Examples of habitual folate intake probability density distributions estimated with AGE MODE for males (*A*) and females (*B*) of several ages in DNFCS-3.

Considering both the mean habitual intakes and EARs as a function of age already gives some insight into folate intake adequacy in the Dutch population (Supplemental Fig. 3). For children, the estimated EAR for adults was interpolated to the requirement for infants to obtain age-dependent EARs. For males, mean intake and requirement were close, whereas in adult women, the mean intake level was substantially lower than the EAR. The estimated proportions of individuals with inadequate intakes have been calculated using the EAR cut-point approach (**Fig. 3**) (16,17). The sharp angle at age 16 y arises because the EAR then reaches it maximum value of 200. Before the age of 16 y, the EAR increases stronger than the intakes, whereas intakes still increase beyond the age of 16 y.

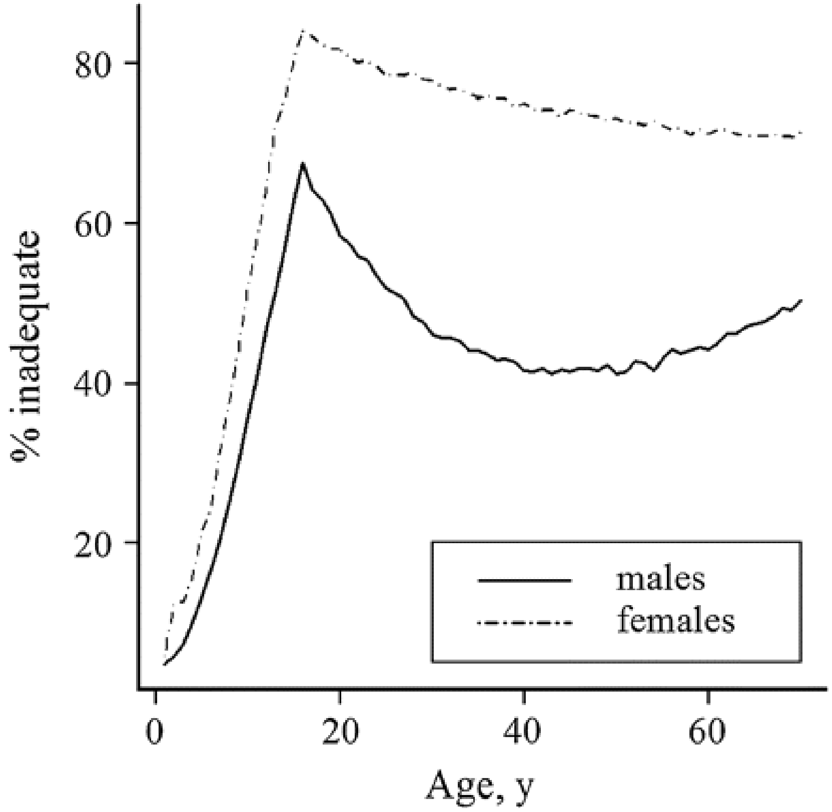

Figure 3 Estimated proportion of individuals with inadequate folate intakes as a function of age for males and females in DNFCS-3 (calculated with the EAR cut-point approach).

Habitual folate intakes estimated with AGE MODE and the ISU method

Direct comparison of estimates from AGE MODE and the method developed by Nusser et al. at ISU (6) is somehow artificial, as estimates by the latter method can only be obtained for population subgroups. A depiction of the mean habitual intake estimates for AGE MODE and the ISU method as a function of age in 1 figure may make a strong appeal for an age-dependent approach (**Fig. 4**). When estimated habitual intake distributions from AGE MODE and the ISU method are compared (Supplemental Fig. 4), they strongly concur. For children, estimated habitual intake distributions from AGE MODE were somewhat wider than those from the method of Nusser et al. (6).

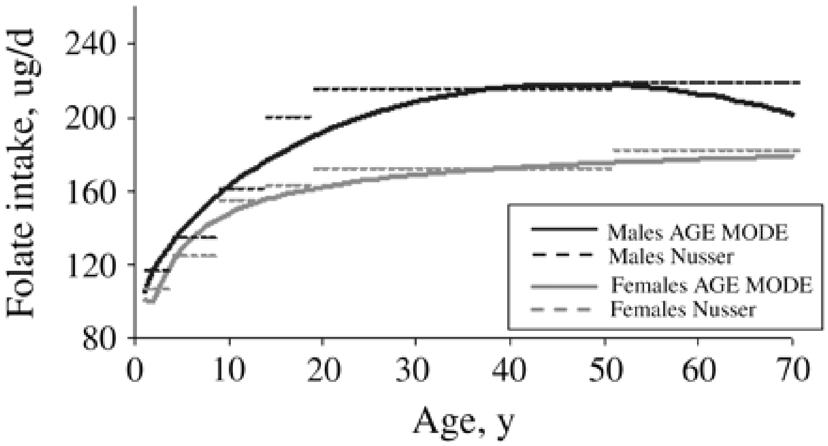

Figure 4 Estimated mean habitual folate intakes from AGE MODE (continuous line) and the ISU method by Nusser et al. (6) (discontinuous line) for males and females.

Discussion

The need for a sound methodology to estimate habitual intakes from short-term dietary measurements, like data from 24-h recalls, is evident. Few methods have been proposed so far, and the method that is most often used, that developed by Nusser et al. at ISU (6), has its limitations. Therefore, we have proposed a new, age-dependent model.

AGE MODE estimates habitual intake distributions as a function of age. Also, subsequent estimates for the prevalence of inadequate intakes of micronutrients can be obtained for any given age. This approach has several advantages above current practice.

Most important may be that with an age-dependent model, it is not necessary to specify subgroups of age. Consequently, variation in intake caused by age is no longer an issue. Another reason to favor an age-dependent model is the fact that intakes of many dietary components are subject to extremely large variations in general. Therefore, estimated habitual intake characteristics may be prone to high uncertainties and be less reliable if numbers are small, for example if smaller subgroups are taken. In AGE MODE, all available data are used to estimate the parameters of the habitual intake distribution. This means that power of precision can be lent from adjacent ages, improving the reliability of the estimates.

In addition, the estimated habitual intake distributions for the individual ages are consistent, whereas transformation parameters can show important differences between adjacent subgroups when estimated individually.

AGE MODE is rather simple and transparent and allows us to gain insight into the underlying data. All steps in the estimation of the habitual intakes are clearly described and illustrated, and the final estimates are depicted in 1 figure with the means of the original observations. The method by Nusser et al. (6), designed to estimate habitual intake distribution for specified subgroups, is operational in statistical software packages developed at ISU (15,18), but the model is extremely complex. The available software generally works well with customary intake data but could be considered a black box, which may be a severe limitation for users, especially in the case of less customary data. It was shown that habitual intake distributions estimated with both methods are comparable.

The fundamental concept of AGE MODE, fitting an age-dependent function, is based on the ideas of Slob, implemented in STEM (19). However, the methodology is completely different. To obtain symmetrically distributed observations, AGE MODE applies the general Box-Cox transformation, whereas in STEM, the log-transformation is used. In our analyses, we often find estimates for λ in the range of 0.20–0.25, significantly different from $\lambda = 0$ (in the case of a natural log transformation).

In the next step, STEM chooses fixed functions with several unknown parameters to fit to the log-transformed data. But because no physical relation between habitual intake and age is known, we propose a more general approach by using fractional polynomial regression. This method simply searches for the best way to describe the data, but has also some drawbacks, like the nonmonotone increase at the lower ages for women. Also important is the back-transformation step. In STEM, the inverse of the forward-transformation is used, which results in estimates for the median, whereas AGE MODE uses Monte Carlo Simulations to obtain estimates for the mean habitual intakes.

Outliers are removed in AGE MODE based on the statistical test proposed by Grubbs and Beck (11). The problem of outliers is complicated. The central question is whether or not to remove outlying observations. On the one hand, they can disturb calculations, but on the other hand, they can contain true information (20).

We inspected the statistical outliers more closely. The lower outlying observations (4 in males, 7 in females) were all due to anomalous low consumption, mainly due to illness. As these observations are not representative for normal intakes, it can be argued to remove them. It should then be considered, however, to remove all individuals who reported lower intake than normal due to illness. High (outlying) folate intakes (8 in males, 7 in females) were mainly due to liver consumption. These high intakes really occur and seem not due to anomalous consumption. One should be aware that leaving out statistical outliers influences (reduces) the intra-individual and also the inter-individual variation. Handling these outliers is a matter of careful judgment. Options are to remove the outliers before transformation and reinclude them afterward, to consider different subgroups (e.g., liver and nonliver consumers), or to include extra covariates in the model.

We proposed a new approach to model the intake of dietary components. Although our methodology is now applicable, it still needs to be further developed. A major issue is the function describing the intakes age-dependently. We used a 2nd order polynomial, as this function appears well capable to describe our data. But the use of higher order polynomials and options to increase the flexibility of the polynomial need to be studied, so the data may be described even better. This is also necessary to obtain consistent results when the age range is extended or restricted.

Furthermore, at this moment, variances are assumed to be the same for all ages. This assumption may not be valid and could be relaxed by estimating inter- and intra-individual variances as a function of age. This will not greatly influence the results but may optimize the model. It should also be investigated how an additional module can be incorporated in AGE MODE to discern between consumers and nonconsumers and to estimate consumption frequency. The methodology for this has already been proposed and is available in S-PLUS (21). Furthermore, as mentioned earlier, the model could be extended with extra covariates. A broader application of AGE MODE may lead to new insights and further improvements.

In conclusion, we proposed an age-dependent methodology to assess the intake of dietary components, named AGE MODE. The model produces estimates for the habitual intake distribution and evaluates intakes in an age-dependent manner. The model may still be further extended, but the feature of age dependency shows clearly described advantages above currently used methods.

ACKNOWLEDGMENTS

The authors thank W. Slob for sharing his work and for his methodological advice.

FOOTNOTES

[1] Supplemental Figures 1–4 are available with the online posting of this paper at jn.nutrition.org.

[4] Abbreviations used: AGE-MODE, age-dependent dietary assessment model; DNFCS-3, third Dutch National Food Consumption Survey; EARs, estimated average requirements; ISU, Iowa State University.

J. Nutr. 136:2916-2920, November 2006

References

1. Biro G, Hulshof KF, Ovesen L, Amorim Cruz JA. Selection of methodology to assess food intake. Eur J Clin Nutr. 2002;56: Suppl 2:S25–32.

2. Willett W. Nature of variation in the diet. In: Willet W., editor. Nutritional epidemiology. New York: Oxford University Press; 1998.

3. Hoffmann K, Boeing H, Dufour A, Volatier JL, Telman J, Virtanen M, Becker W, De Henauw S. Estimating the distribution of usual dietary intake by short-term measurements. Eur J Clin Nutr. 2002;56: Suppl 2:S53–62.

4. Buck RJ, Hammerstrom KA, Ryan PB. Estimating long-term exposures from short-term measurements. J Expo Anal Environ Epidmiol. 1995;5:359–73.

5. Gay C. Estimation of population distributions of habitual nutrient intake based on a short-run weighed food diary. Br J Nutr. 2000;83:287–93.

6. 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.

7. Slob W. Modeling long-term exposure of the whole population to chemicals in food. Risk Anal. 1993;13:525–30.

8. Anonymous. S-PLUS 7.0, Guide to statistics. Vol 1 and 2. Seattle: Insightful Corporation; 2005.

9. Royston P, Altman DG. Approximating statistical functions by using fractional polynomial regression. Statistician. 1997;46:411–22.

10. Grubbs F. Procedures for detecting outlying observations in samples. Technometrics. 1969;11:1–21.

11. Grubbs FE, Beck G. Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics. 1972;14:847–54.

12. Millard SP. Environmental stats for S-PLUS. Seattle: Insightful Corporation; 2002.

13. Voedingscentrum. Zo eet Nederland. Results of the Dutch Food Consumption Survey 1997/1998. Den Haag: Voedingscentrum; 1998.

14. Voedingscentrum. NEVO-tabel 2001. Den Haag: Voedingscentrum; 2001.

15. Iowa State University. A user's guide to C-SIDE. Ames (IA): Iowa State University, Department of Statistics and Center for Agricultural and Rural Development; 1996.

16. Institute of Medicine. Dietary reference intakes: applications in dietary assessment. Washington: National Academy Press; 2000.

17. Carriquiry AL. Assessing the prevalence of nutrient inadequacy. Public Health Nutr. 1999;2:23–33.

18. Iowa State University. A user's guide to SIDE. Ames (IA): Iowa State University, Department of Statistics and Center for Agricultural and Rural Development; 1996.

19. Slob W. Modeling human exposure to chemicals in food. Bilthoven: RIVM; 1993.

20. Bakker MI, Slob W. Omgaan met uitbijters in innameberekeningen. Bilthoven: RIVM; 2005.

21. Slob W, Bakker MI. Probabilistische berekening van inname van stoffen via incidenteel geconsumeerde voedingsproducten. Bilthoven: RIVM; 2004.