Estimation of the Usual Intake Distributions of Ratios of Dietary Components

Dietary Assessment Research Series Report 5

A.L. Carriquiry, W.A. Fuller, J.J. Goyeneche, and K.W. Dodd

Staff Report 95-SR 79
December 1995

Estimation of the Usual Intake Distributions of Ratios of Dietary Components

Dietary Assessment Research Series Report 5

A.L. Carriquiry, W.A. Fuller, J.J. Goyeneche, and K.W. Dodd

Staff Report 95-SR 79
December 1995

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070

A.L. Carriquiry is associate professor, W.A. Fuller is distinguished professor, J.J. Goyeneche is a research assistant, and K.W. Dodd is a research assistant, in the statistics department, Iowa State University.

Partial funding support was provided under Cooperative Research Agreement #58-3198-2-006 between the Agricultural Research Service, USDA, and Iowa State University.

The authors gratefully acknowledge comments and suggestions received from Dr. Phillip Kott from USDA/NAS, Dr. Patricia Guenther from USDA/ARS, and from Dr. Helen Jensen from the Center for Agricultural and Rural Development, Iowa State University.

ABSTRACT

In recent years, it has become apparent that ratios of usual intakes of dietary components, such as the percent of calories from fat or the percent of calories from saturated fat, are important indicators of dietary adequacy. The problem of estimating a percentile of the distribution of ratios of usual intakes of nutrients presents problems that arise from the fact that, from a statistical point of view, ratios of dietary components can be viewed as ratios of random variables. The methods developed by Nusser et al. (1995) to estimate usual intake distributions of nutrients, for example, cannot be directly applied to ratios, since the ratio of usual intakes is not equal to the mean of the ratio of daily intakes.

In this paper we discuss the problem of estimating the intake distribution of ratios of dietary components by using an extension of the method developed by Beale (1962). The method we propose consists of first estimating the daily ratio of dietary components for each individual to obtain a nearly unbiased ratios estimate, and then applying the method developed at Iowa State University to estimate the distribution of usual intakes of the ratio.

We use data collected in the 1989-91 CSFII, and consider six ratios to illustrate the procedure.

ESTIMATION OF THE USUAL INTAKE DISTRIBUTIONS OF RATIOS OF DIETARY COMPONENTS

The Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA) conducts periodic dietary surveys for assessing the dietary status of the population. One important objective is to estimate the proportion of the population (or of different subpopulations) who may be consuming inadequate or excessive amounts of any of a set of nutrients. This information can then be used to tailor government policy for public health and food assistance programs, and to plan intervention in food and nutrition programs such as school lunches.

The problem of estimating the percent of the population with inadequate or excessive nutrient intake can be formulated, in statistical terms, as the problem of estimating a percentile of the usual intake distribution for the nutrient. Here, *usual intake* is defined as the long-run average intake of a nutrient, and the *distribution of usual intakes* is defined as the distribution of those long-run averages for the population of interest. A method for estimating usual intake distributions for nutrients was proposed by Nusser et al. (1995). When dietary data are collected on consecutive days, then observations for an individual cannot be assumed to be independent, and it becomes necessary to estimate the day-to-day correlation for nutrient intake. A procedure to estimate day-to-day correlations of nutrient intake was proposed by An and Carriquiry (1992) and by Carriquiry et al. (1995).

In recent years, however, it has become apparent that *ratios of usual intakes of dietary components*, such as the percent calories from fat or the percent calories from saturated fat, are also important indicators of dietary adequacy. Initiatives such as Healthy People 2000 (U.S. Department of Health and Human Services, USDHHS, 1991) explicitly refer to several of these ratios, and establish goals to be reached by the year 2000. Two such goals are to lower the intake of fat and saturated fat in the population to 30% and 10% of calories, respectively. In principle, the problem of assessing the status of the population regarding their intake of ratios of nutrients is similar to the one described earlier. Ratios of random variables, however, present several problems from a statistical point of view, and methods presented by Nusser et al. (1995), and by An and Carriquiry (1992) and Carriquiry et al. (1995) cannot be directly applied. This is because the ratio of usual intakes is not equal to the mean of the ratios of daily intakes.

Ideally, a method for estimating the usual intake distribution of a ratio of dietary components would need to be based on the bivariate distribution of the numerator and the denominator in the ratio. If a bivariate transformation to jointly map the numerator and the denominator into bivariate normality were available, then the usual intake distribution of the ratio could be estimated by extending Nusser et al.'s (1995) procedure to the multivariate case. Usual intakes of the numerator and the denominator in the bivariate normal scale could be obtained by fitting a bivariate measurement error model that accounts for the correlation between numerator and denominator. The adjusted inverse bivariate transformation would then be used to jointly map the estimated usual intakes into the original bivariate scale. The estimated usual intake of the ratio in the original scale would then be obtained by applying the theory for estimation of distributions of transformed random variables, and then deriving the marginal distribution of the ratio. A method for obtaining the distribution of transformed random variables is described, for example, in Lindgren (1976). A procedure similar to the one described is currently under development. The method and results presented here, however, are an approximation to the multivariate extension to the ISU method.

In this report, we discuss the problem of estimating the correlation among intake days and the usual intake distributions for ratios of dietary components that are estimated using an extension of the method presented by Beale (1962). We use data collected in the 1989, 1990, and 1991 Continuing Survey of Food Intake by Individuals (CSFII), and consider six ratios: percent calories from total fat, monounsaturated fat, polyunsaturated fat, saturated fat, carbohydrates, and protein. The methods presented here can be applied to any other ratio of dietary components. A nearly unbiased estimate of the intake of the ratio for each individual in the group and each intake day is obtained. Correlations (using the ratio estimates) are estimated for each of five age-sex categories: men 20 to 59 years of age, men 60 years old and older, women 20 to 59 years old, women 60 years of age and older, and individuals 19 years old and younger. Correlations were estimated following the two-step approach presented in Carriquiry et al. (1995). Estimated usual intake distributions were obtained using the ISU method described in Nusser et al. (1995).

Data and Data Adjustments

Data used to estimate the correlation among intake days for ratios of dietary components were obtained from the combined 1989-91 CSFII. Data were divided into five age-sex categories, and a correlation estimate was obtained for each ratio in each age-sex group. There were 2,494 men between 20 and 59 years of age and 887 mean over 59 years of age, 3,168 non pregnant, non lactating women ages 20 to 59, 1,453 women over 59 years of age, and 3,710 individuals younger than 20 years of age. Intake data for each individual were collected on three consecutive days, and thus observations for an individual are not independent.

Dietary intake data were examined to determine the effect of survey-related factors on intake. Following Nusser et al. (1995) intake data were ratio-adjusted to remove the effect of day of the week, and interview sequence (confounded with month of the year) on mean intake. The ratio-adjustment produced observations for each of the three days with the mean adjusted to be the same as the mean intake observed for the first day. Similarly, the variance of the observed intakes on days two and three was also adjusted to equal the first day variance.

Prior to computing ratios of dietary components, intake data for fat, protein, carbohydrates, and alcohol were converted from their original units into calories by following the recommendations in the National Academy of Sciences report (1989). Conversion factors used were 4 kcal/g, 9 kcal/g, and 7 kcal/g for protein, fat, and alcohol, respectively. Intake data expressed in calories were examined for internal consistency. For example, it was observed that, for several individuals in the sample, the sum of calories from saturated fat, polyunsaturated fat, and monounsaturated fat exceeded the amount of calories from total fat. Similarly, it was observed that for some individuals, the sum of calories from fat, carbohydrates, protein, and alcohol exceeded total calories consumed. Since these apparent inconsistencies have implications when considering ratios of dietary components, data were edited prior to analysis. We applied two simple corrections:

1. If the sum of calories from saturated fat, monounsaturated fat, and polyunsaturated fat exceeded the calories from total fat, then total fat was set equal to the sum of the components.

2. Total calories consumed was set equal to the sum of calories calculated from fat, carbohydrates, protein, and alcohol.

In addition, and to avoid computational difficulties, total calories consumed was set to be equal to 0.01 whenever the datum for an individual was equal to zero. The intake data set, even after corrections, included several individuals with intake equal to zero for some or all interview days and all dietary components. For these individuals, total calories consumed was set to 0.01, but no other intakes were modified. The proportion of individuals with zero intakes for some or all intake days was highest in the sex-age group corresponding to persons 0 to 19 years of age, and females 20 to 59 years of age. There were no males in any age group with zero consumption of all dietary components on any day.

A Nearly Unbiased Ratio Estimate

A naive estimate for the ratio of two dietary components for an individual on any day is given by Y_{ij}/X_{ij}, where Y_{ij} is the adjusted observed intake of, for example, calories from total fat, for the i-th individual on the j-th day, and X_{ij} is the corresponding intake of calories for that individual on that day. Let Y and X be unbiased estimates of the individual's usual intake of the components, i.e., $E(Y_{ij}|i) = y_i$ and $E(X_{ij}|i) = x_i$. It is well known that the ratio Y/X is a biased estimate of y/x. Expanding the ratio $X^{-1}Y$ in a Taylor series around $y/x = R$,

$$\frac{Y}{X} \doteq R + \frac{1}{x}(Y - RX) - \frac{1}{x^2}[(X-x)(Y-y) - R(X-x)^2]. \qquad (1)$$

It follows that an approximate expression for the expected value of the ratio Y/X is

$$E(\frac{Y}{X}) \doteq R - \frac{1}{x^2}(\sigma_{XY} - R\sigma_X^2), \qquad (2)$$

where $\sigma_X^2 = E\{(X-x)^2\}$ and $\sigma_{XY} = E\{(X-x)(Y-y)\}$. In expression (2), σ_X^2 is the variance of the denominator, and σ_{XY} is the covariance between the numerator and the denominator in the ratio.

The naive ratio estimator for each individual in the sample is obtained by computing the ratio of the observed intake means. Letting $3^{-1}\bar{Y}_i = \sum_{j=1}^{3} Y_{ij}$ and $3^{-1}\bar{X}_i = \sum_{j=1}^{3} X_{ij}$ represent the observed intake means for the numerator and the denominator, respectively, the naive ratio estimator is

$$\hat{R}_i = \frac{\bar{Y}_i}{\bar{X}_i}, \tag{3}$$

with approximate bias equal to

$$E(\hat{R}_i) - R_i = -x_i^{-2} 3^{-1} (\sigma_{i,X}^2 - R_i \sigma_{i,XY}), \tag{4}$$

where $\sigma_{i,X}^2$ is the within individual variance for the denominator, and $\sigma_{i,XY}$ is the within person covariance between the numerator and the denominator.

A less biased estimator for the ratio was proposed by Beale (1962) and later discussed by Cochran (1977, p. 176), and is given by

$$\ddot{R}_i = (\bar{X}_i^2 + 3^{-1}\hat{\sigma}_{i,X}^2)^{-1}(\bar{X}_i\bar{Y}_i + 3^{-1}\hat{\sigma}_{i,XY}), \tag{5}$$

where $\hat{\sigma}_{i,X}^2$ and $\hat{\sigma}_{i,XY}$ are estimators of $\sigma_{i,X}^2$ and $\sigma_{i,XY}$, respectively. An expression for the bias of the estimator proposed by Beale (1962) is given in the Appendix.

Notice that expression (5) provides a less biased estimator for the ratio of usual intakes for an individual. In our work, we are interested in estimating the distributions of the ratios of usual intakes. Therefore, we extend the method proposed by Beale (1962) to obtain $\tilde{R}_{ij}$, a less biased estimator for a ratio for the i-th individual on the j-th day. To compute $\tilde{R}_{ij}$, we need estimates of $\sigma_{i,X}^2$ and of $\sigma_{i,XY}$. The direct estimates of these parameters are computed on only two degrees of freedom for each individual. In order to obtain estimates of $\sigma_{i,X}^2$ and of $\sigma_{i,XY}$ with less variance than the direct estimates, we develop a model for the variances and covariances. These model estimates are smoother than the direct estimates.

Refer to expression (5), and consider the problem of estimating $\sigma_{i,X}^2$ and $\sigma_{i,XY}$. Let

$$\begin{aligned} s_{i,X}^2 &= \frac{1}{2}\sum_{j=1}^{3}(X_{ij} - \bar{X}_i)^2, \\ s_{i,XY} &= \frac{1}{2}\sum_{j=1}^{3}(X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i), \end{aligned}$$

denote estimators of the within person variance for the denominator and the within person covariance between the numerator and the denominator, respectively. Furthermore, let

$$b_{YX} = (\sum_{i=1}^{n}(s_{i,X}^2)^2)^{-1}\sum_{i=1}^{n} s_{i,XY} \tag{6}$$

denote the ratio of the within person covariances between the numerator and the denominator to the within person variances for the denominator. Smooth estimates of the within individual variance for the denominator, and for the covariance between numerator and denominator can be obtained as is described next. Let

$$h_i = \bar{Z}_i^C, \tag{7}$$

where

$$\bar{Z}_i = 3^{-1}\sum_{j=1}^{3}(X_{ij} + 0.02\bar{X}_{..})^a, \tag{8}$$

$C = 2(a^{-1} - 1)$, and a is the estimated best power for the denominator (obtained as is described, for example, in Nusser et al. (1995)). In expression (8), $\bar{X}_{..}$ represents the grand mean (over all individuals) of the denominator of the ratio. In (8), the grand mean is added to the observations just to move the data away from zero. Note that h_i is proportional to the derivative of $\bar{Z}_i$, and is thus an approximation to the variance of the individual mean for the denominator when data are power-transformed. If the X_{ij}^a have homogeneous variances, then the variance of $\bar{X}_i$ is proportional to h_i. If this relationship holds, then estimates for the within individual variances that make use of the whole sample (rather than the three days available for each individual) can be obtained. An improved estimate of the within person variance for the denominator can be obtained from the regression of $s_{i,X}^2$ on h_i. Let $s_{i,X}^2 = g_i$. Then, an estimator of $\sigma_{i,X}^2$ is

$$\hat{g}_i = \hat{\gamma}_0 + \hat{\gamma}_1 h_i, \tag{9}$$

where $(\hat{\gamma}_0, \hat{\gamma}_1)$ are the regression coefficients in the regression of g_i on $(1, h_i)$. The estimator $\hat{\gamma}_0$ is restricted to be nonnegative. If in estimating the regression coefficients in (9) the intercept $\hat{\gamma}_0$ is negative, then we set $\hat{\gamma}_0 = 0$ and compute the regression through the origin. The $\hat{g}_i$ of (9) is an estimator of $\sigma^2_{i,X}$, and $b_{YX}\hat{g}_i$ is an estimator of $\sigma_{i,XY}$. These estimated within person variances and covariances can be substituted in Beale's expression for the ratio estimator as given in (5), to obtain

$$\begin{aligned} \ddot{R}_i &= (\bar{X}_i^2 + 3^{-1}\hat{g}_i)^{-1}(\bar{X}_i\bar{Y}_i + 3^{-1}b_{YX}\hat{g}_i) \\ &= (\bar{X}_i + 3^{-1}\bar{X}_i^{-1}\hat{g}_i)^{-1}(\bar{Y}_i + 3^{-1}\bar{X}_i^{-1}b_{YX}\hat{g}_i). \end{aligned} \tag{10}$$

In estimating the distribution of R_i we use $\ddot{R}_i$ as a nearly unbiased estimator of R_i. That is, $\ddot{R}_i$ is a nearly unbiased estimator of the mean intake of the ratio. In using the program for the usual intake distribution it is necessary to estimate the within and between person variance components of $\ddot{R}_i$. We therefore need to create three observations $\tilde{R}_{ij}$ for the ratio for each individual on each day, such that the within individual mean of $\ddot{R}_i$ equals $\ddot{R}_i$.

The Taylor expansion of the error in (3) is

$$\hat{R}_i - R_i \doteq x_i^{-1}(\bar{Y}_i - R_i\bar{X}_i), \tag{11}$$

where $R_i = y_i x_i^{-1}$. This suggests that three "observations" can be created on the ratio estimate as

$$\tilde{R}_{ij} = \ddot{R}_i + (\bar{X}_i + 3^{-1}\bar{X}_i^{-1}\hat{g}_i)^{-1}(Y_{ij} - \hat{R}_i X_{ij}), \tag{12}$$

for $j = 1, 2, 3$, where $\ddot{R}_i$ is defined in (10) and $\hat{R}_i$ is as given in (3). These new observations have the following properties:

- The mean of the estimated ratios for the i-th individual is $\ddot{R}_i$. This is easily shown by observing that

$$\begin{aligned} 3^{-1}\sum_{j=1}^{3}\tilde{R}_{ij} = \ddot{R}_i + (\bar{X}_i + 3^{-1}\bar{X}_i^{-1}\hat{g}_i)^{-1}3^{-1}\sum_{j=1}^{3}(Y_{ij} - \hat{R}_i X_{ij}) \\ = \ddot{R}_i. \end{aligned}$$

Table 1. Means and standard deviations for estimated ratios $\tilde{R}_{ij}$ of dietary components in each sex-age group.

Calories Source	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Total fat	0.36	0.34	0.35	0.33	0.35
	(0.065)	(0.070)	(0.070)	(0.070)	(0.056)
Monounsaturated fat	0.13	0.13	0.13	0.12	0.13
	(0.028)	(0.030	(0.030)	(0.029)	(0.024)
Saturated fat	0.12	0.12	0.12	0.11	0.13
	(0.029)	(0.031)	(0.031)	(0.031)	(0.029)
Polyunsaturated fat	0.07	0.07	0.07	0.07	0.06
	(0.022)	(0.023)	(0.024)	(0.025)	(0.021)
Carbohydrates	0.46	0.48	0.48	0.50	0.50
	(0.077)	(0.080)	(0.083)	(0.077)	(0.068)
Protein	0.16	0.17	0.16	0.17	0.15
	(0.036)	(0.037)	(0.039)	(0.039)	(0.031)

- The within individual variance for the estimated ratio $\tilde{R}_{ij}$ divided by three is an estimator of the Taylor approximation to the variance of $\ddot{R}_i$ for the i-th individual.

The estimated ratios $\tilde{R}_{ij}$, $j = 1, 2, 3$ for individual i can now be used to obtain the correlation among intake days for each ratio and each age-sex group, and to estimate the usual intake distributions of the ratios. Sample statistics (means and standard deviations) for the estimated ratios $\tilde{R}_{ij}$ for all age-sex groups are presented in Table 1. Sample statistics for the naive ratio estimates $\hat{R}_{ij}$ were very similar to those given in Table 1, and are not presented here. In Figure 1 we present plots of $\hat{R}_{ij}$ versus $\tilde{R}_{ij}$ for females 20 to 59 years of age and persons 0 to 19 years of age, for the ratios corresponding to total fat and carbohydrates.

Estimation of Correlations and Usual Intake Distributions

Consider the observations $\tilde{R}_{ij}$ constructed for each individual in an age-sex group, on each of three days. These observations can now be treated as the usual univariate observations on intake of a dietary component, and methods proposed in Carriquiry et al. (1995) and in Nusser et al. (1995) for estimation of day-to-day correlations and usual intake distributions can be applied.

Smooth estimates of the day-to-day correlations were computed following the two-step procedure outlined in Carriquiry et al. (1995). The correlations were estimated jointly for 27 dietary

Table 2. Smooth day-to-day correlation estimates for estimated ratios $\tilde{R}_{ij}$ of dietary components in each sex-age group

Components	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Total fat	0.082	0.079	0.055	0.000	0.035
Monounsaturated fat	0.073	0.071	0.035	0.000	0.007
Saturated fat	0.112	0.087	0.119	0.045	0.131
Polyunsaturated fat	0.084	0.097	0.061	0.031	0.034
Carbohydrates	0.091	0.071	0.074	0.000	0.068
Protein	0.082	0.136	0.063	0.103	0.017

components, water, and the six ratios, so that 34 direct estimates of correlation were available for each age-sex group. In the first step, direct correlation estimates were obtained as is suggested in An and Carriquiry (1992). A two-factor model was then fit to the direct correlation estimates to obtain a set of smooth estimates of the day-to-day correlations, as is discussed in Carriquiry et al. (1995). Those smooth estimates are presented in Table 2, for all ratios and all age-sex groups.

To assess the percentage of the population with insufficient (or excessive) ratio, we used the estimated $\tilde{R}_{ij}$ to obtain usual intake distributions. We followed the procedure outlined in Nusser et al. (1995): within each age-sex group, and for each ratio, data were first adjusted to account for survey-related and other effects. The adjustment consisted of two steps. First, the mean intake for the second and third days was adjusted to the mean intake observed on the first day. In the second step, the variance of the distribution of intakes on days two and three was adjusted to be equal to the first day variance. Prior to adjustment, data were transformed using a nonlinear power transformation using a method similar to the one outlined in Lin and Vonesh (1989). Three day weights computed for the 89, 90, and 91 CSFII were incorporated into all the procedures.

The correlation estimates presented in Table 2 were used in the software for estimation of usual intake distributions. The ISU method (Nusser et al., 1995) was used to estimate usual intake distributions for the estimated ratios $\tilde{R}_{ij}$ and for the naive ratio estimates $\hat{R}_{ij}$ for each age-sex category. Figure 2 shows the estimated usual intake distributions for the ratio of calories from total fat to total calories for males ages 20 to 59 (figure (a)) and females ages 20 - 59 (figure (c)), and for the ratio of calories from saturated fat to total calories (figures (b) and (d) for males and females 60 years of age and older, respectively). Vertical lines drawn on the plots indicate the

Table 3. Estimated percentiles of the usual intake distribution of $\tilde{R}_{ij}$ for males, 20 to 59 years of age

Percentile	T. fat	M. fat	S. fat	P. fat	Carb.	Prot.
0.01	0.230	0.082	0.072	0.040	0.313	0.110
0.05	0.276	0.099	0.088	0.047	0.359	0.124
0.10	0.297	0.108	0.097	0.051	0.382	0.132
0.50	0.359	0.135	0.124	0.068	0.457	0.162
0.90	0.416	0.161	0.154	0.089	0.534	0.197
0.95	0.435	0.170	0.164	0.096	0.559	0.207
0.99	0.473	0.189	0.183	0.110	0.606	0.228

Year 2000 Objectives for consumption of calories from fat and saturated fat.

Selected estimated percentiles of the usual intake distributions of each ratio are presented in Tables 3, 4, 5, 6, 7, for each of the age-sex groups, respectively. . Percentiles computed from the estimated usual intake distributions that were obtained using the naive ratios $\hat{R}_{ij}$ were very similar, and are not presented here.

Year 2000 Objectives (USDHHS 1991) establish the following goals: individuals in any sex-age group should consume a diet that contains no more than 30% of calories from fat, and no more than 10% of calories from saturated fat. Ideally, the proportion of the population with consumption of fat and saturated fat that is high relative to the rest of their diet, would be reduced by the year 2000. To examine the current status of the five subpopulations in this study in relation to the Year 2000 Objectives, we computed the percent of each subpopulation with consumption of calories from fat above 30% and consumption of calories from saturated fat exceeding 10% of total calories. These percentages are presented in Table 8. In the first two columns, we show the results that were obtained when usual intake distributions were estimated using the Beale ratio estimator $\tilde{R}_{ij}$. The last two columns show results obtained from the naive ratio estimator $\hat{R}_{ij}$.

From Table 8 it appears that a large proportion of the population consumes well above the recommended amount of fat and saturated fat relative to the rest of their diet. When estimating percentiles of the tail of the usual intake distributions, results obtained from the usual intake distribution of the Beale ratio estimates (columns 1 and 2) differ from those that are obtained from the distribution of the naive ratio estimator (columns 3 and 4). The Beale procedure produces a larger estimated proportion exceeding the goals.

Table 4. Estimated percentiles of the usual intake distribution of $\tilde{R}_{ij}$ for males,60 years of age and older

Percentile	T. fat	M. fat	S. fat	P. fat	Carb.	Prot.
0.01	0.213	0.074	0.062	0.039	0.307	0.110
0.05	0.249	0.089	0.075	0.046	0.357	0.124
0.10	0.269	0.097	0.083	0.051	0.383	0.132
0.50	0.343	0.127	0.116	0.069	0.474	0.166
0.90	0.414	0.159	0.151	0.092	0.560	0.208
0.95	0.435	0.168	0.162	0.099	0.584	0.221
0.99	0.475	0.187	0.186	0.114	0.628	0.249

Table 5. Estimated percentiles of the usual intake distribution of $\tilde{R}_{ij}$ for females, 20 to 59 years of age

Percentile	T. fat	M. fat	S. fat	P. fat	Carb.	Prot.
0.01	0.215	0.074	0.066	0.039	0.347	0.104
0.05	0.255	0.091	0.082	0.047	0.386	0.121
0.10	0.276	0.100	0.090	0.052	0.406	0.130
0.50	0.348	0.128	0.119	0.070	0.480	0.159
0.90	0.408	0.154	0.148	0.092	0.563	0.194
0.95	0.424	0.162	0.158	0.099	0.588	0.207
0.99	0.456	0.178	0.176	0.114	0.634	0.237

Table 6. Estimated percentiles of the usual intake distribution of $\tilde{R}_{ij}$ for females, 60 years of age and older

Percentile	T. fat	M. fat	S. fat	P. fat	Carb.	Prot.
0.01	0.194	0.067	0.055	0.035	0.349	0.108
0.05	0.233	0.082	0.071	0.044	0.393	0.122
0.10	0.255	0.090	0.079	0.049	0.417	0.130
0.50	0.331	0.121	0.111	0.068	0.500	0.162
0.90	0.403	0.152	0.149	0.093	0.581	0.203
0.95	0.423	0.162	0.160	0.102	0.604	0.216
0.99	0.462	0.180	0.182	0.122	0.648	0.242

Table 7. Estimated percentiles of the usual intake distribution of $\tilde{R}_{ij}$ for persons, 0 to 19 years of age

Percentile	T. fat	M. fat	S. fat	P. fat	Carb.	Prot.
0.01	0.227	0.075	0.062	0.032	0.389	0.087
0.05	0.271	0.093	0.075	0.039	0.426	0.105
0.10	0.290	0.101	0.083	0.043	0.443	0.115
0.50	0.344	0.127	0.131	0.058	0.504	0.148
0.90	0.400	0.150	0.161	0.078	0.572	0.180
0.95	0.416	0.158	0.171	0.087	0.595	0.190
0.99	0.448	0.171	0.190	0.113	0.655	0.210

Table 8. Estimated percentage of the population with intakes of calories from total fat above 30 percent and intakes of calories from saturated fat above 10 percent

	Beale		Naive	
Group	Tot. fat	Sat. fat	Tot. fat	Sat. fat
Males 20-59	0.891	0.873	0.882	0.872
Males 60 and older	0.773	0.726	0.729	0.694
Females 20-59	0.810	0.807	0.776	0.786
Females 60 and older	0.701	0.666	0.665	0.644
Persons 0-19	0.853	0.930	0.775	0.906

Note: First two columns: Beale estimator. Second two columns: naive ratio estimator

Conclusions

Assessing the ratios of consumption of dietary components, such as the percent calories from fat or from saturated fat, is recognized as an important aspect of dietary assessment for populations. The U.S. Department of Health and Human Services has explicitly established goals to be reached by the year 2000. Under those goals, no more than 30% of all calories consumed should come from fat, and no more than 10% should come from saturated fat.

From a statistical viewpoint, ratios of random variables present unique problems, and cannot, in general, be analyzed as univariate random variables. In particular, it is well known that, in general, the ratio of the expected value of two random variables is not equal to the expected value of the ratio. In light of this, methods presented by An and Carriquiry (1992), and Carriquiry et al. (1995) for obtaining smooth estimates of day-to-day correlations for dietary components, and the ISU method for estimating usual intake distributions (Nusser et al., 1995) cannot be directly applied to the ratio of two components.

Two possible solutions to the problem of estimating usual intake distributions for a ratio of dietary components consist of the following: (1) Derive an multivariate extension of the ISU method, and then obtain the distribution of the ratio as the marginal estimated distribution in the original scale, or (2) Apply the univariate procedure to an unbiased (or nearly unbiased) estimate of the ratio obtained for each individual on each day. We are currently developing a bivariate extension to the ISU method for estimating usual intake distributions. Results from this work will be presented in a later report. In this preliminary report, we present a method for estimating the intake ratio of two dietary components for an individual on any day, and use the estimated ratios to obtain usual intake distributions. We used data collected in the 1989-90-91 CSFII, and considered five sex-age groups: males 20-59 and older than 59, females 20-59 and older than 59, and persons 0 to 19 years of age. The procedure we propose for obtaining a nearly unbiased ratio estimate is an extension of the method proposed by Beale (1962). Beale's ratio estimator, as many other estimators in the literature, relies on asymptotic arguments. As was pointed out by Kott (1995, personal communication), this may create problems when the method is applied to individuals with as little as two observations. However, the estimator presented by Beale (1962) has nice small sample properties, in the sense that regardless of sample size, the denominator is guaranteed to be at least as large as the numerator, and thus the ratio does not "blow up". Results are compared to those that would be obtained by using the naive ratio estimator to compute usual intake distributions.

Further Studies

In the analysis of this paper we did not restrict the sum of the Beale estimated ratios to be one. A simple modification such as, for each person, dividing the Beale ratio for a component by the sum of the Beale ratios for all components would produce estimated ratios that sum to one.

The presence of days of zero consumption introduce difficulties into the analysis of ratios. We chose to replace the zero total consumption by a small positive number. The effect of this operation deserves further study. A related issue is the editing of the data. We modified some data to produce internally consisten estimates. Alternative editing rules could be investigated.

Appendix A. Expansion of Beale's Estimator

Let $(\bar{Y}, \bar{X})$ be the sample means and let $s_{\bar{X},\bar{Y}}$ and $s^2_{\bar{X}}$ be sample estimates of the covariance between $\bar{Y}$ and $\bar{X}$, and of the variance of $\bar{X}$, respectively. Then, Beale's estimator of the ratio $R = x^{-1}y$, where $(x, y) = E(X, Y)$ is

$$\ddot{R} = \frac{\bar{Y} + \bar{X}^{-1} s_{\bar{X}\bar{Y}}}{\bar{X} + \bar{X}^{-1} s^2_{\bar{X}}}. \tag{1}$$

We write the error in $\ddot{R}$ as

$$\ddot{R} - R = \frac{\bar{Y} - R\bar{X} + \bar{X}^{-1}(s_{\bar{X}\bar{Y}} - Rs^2_{\bar{X}})}{x(1+\delta)}, \tag{2}$$

where

$$\delta = x^{-1}(\bar{X} - x + \bar{X}^{-1} s_{\bar{X}^2}). \tag{3}$$

We assume δ is small, and that $x^{-1}(\bar{Y} - R\bar{X} + \bar{X}^{-1}(s_{\bar{X}\bar{Y}} - Rs^2_{\bar{X}}))$ is also small. In the usual sampling situation, $(\bar{X} - x) = Op(n^{-\frac{1}{2}})$ and $s^2_{\bar{X}} = Op(n^{-1})$, where n is the sample size. Then, expanding the denominator in δ we have

$$\ddot{R} - R = x^{-1}[\bar{Y} - R\bar{X} + \bar{X}^{-1}(s_{\bar{X}\bar{Y}} - Rs^2_{\bar{X}})][1 - \delta] + : \text{Remainder} \tag{4}$$

where the remainder is of smaller order than the included terms. Then

$$\begin{aligned} \ddot{R} - R \doteq\ & x^{-1}[\bar{Y} - R\bar{X} + x^{-1}(s_{\bar{X}\bar{Y}} - Rs^2_{\bar{X}})] - x^{-2}[(\bar{Y} - R\bar{X})(\bar{X} - x) \\ & + (\bar{Y} - R\bar{X})x^{-1}s^2_{\bar{X}} + x^{-1}(s_{\bar{X}\bar{Y}} - Rs^2_{\bar{X}})(\bar{X} - x) \\ & + x^{-2}s^2_{\bar{X}}(s_{\bar{X}\bar{Y}} - Rs^2_{\bar{X}})] + \text{Remainder}. \end{aligned}$$

If we adopt the convention that $\sigma^2_{\bar{X}} = Op(n^{-1})$, and assume that

$$(s^2_{\bar{X}}, s_{\bar{X}\bar{Y}}) = (\sigma^2_{\bar{X}}, \sigma_{\bar{X}\bar{Y}} + Op(n^{-\frac{3}{2}}), \tag{5}$$

and further assume that the ratio is such that we can approximate the expectation with the expectation of the leading term in the expansion, then

$$
\begin{aligned}
E\{\ddot{R} - R\} &= x^{-2}(\sigma_{\bar{X}\bar{Y}} - R\sigma^2_{\bar{X}}) - x^{-2}(\sigma_{\bar{X}\bar{Y}} - R\sigma^2_{\bar{X}}) + O(n^{-2}) \\
&= O(n^{-2}).
\end{aligned}
$$

Thus, the bias of Beale's estimator is $O(n^{-2})$, while the bias of the simple estimator is $O(n^{-1})$.

REFERENCES

An, A. and A. L. Carriquiry. 1992. Estimation of the correlation among days for transformed nutrient intake. Report prepared for the Agricultural Research Service (formerly Human Nutrition Information Service), U.S. Department of Agriculture. Department of Statistics, Iowa State University.

Beale, E. M. L. 1962. Some uses of computers in operational research. *Industrielle Organisation* **31**, 27-28.

Carriquiry, A. L., W. A. Fuller, J. J. Goyeneche, and H. H. Jensen. 1995. Estimated correlations among days for the combined 1989-91 CSFII. Preliminary report prepared for the Agricultural Research Service, U.S. Department of Agriculture. Department of Statistics, Iowa State University.

Cochran, W. G. 1977. *Sampling Techniques.* 3rd. ed., Wiley, New York.

Lin, L. I-K, and E. F. Vonesh. 1989. An empirical nonlinear data-fitting approach for transforming data to normality. *The American Statistician* **43**, 237-243.

Lindgren, B. W. 1976. *Statistical Theory.* 3rd. ed., MacMillan, New York.

National Academy of Sciences, Food Nutrition Board. 1989. Recommended Dietary Allowances. National Academy Press, 10th ed. Washington, D.C.

National Health Promotion and Disease Prevention Objectives: Healthy People 2000. 1991. Public Health Service, U.S. Department of Health and Human Services Publication No. (PHS) 91-50212.

Nusser, S. M., A. L. Carriquiry, W. A. Fuller, and K. W. Dodd. 1995. A semiparametric approach to estimating usual intake distributions. Tentatively accepted by the *Journal of the American Statistical Association.*

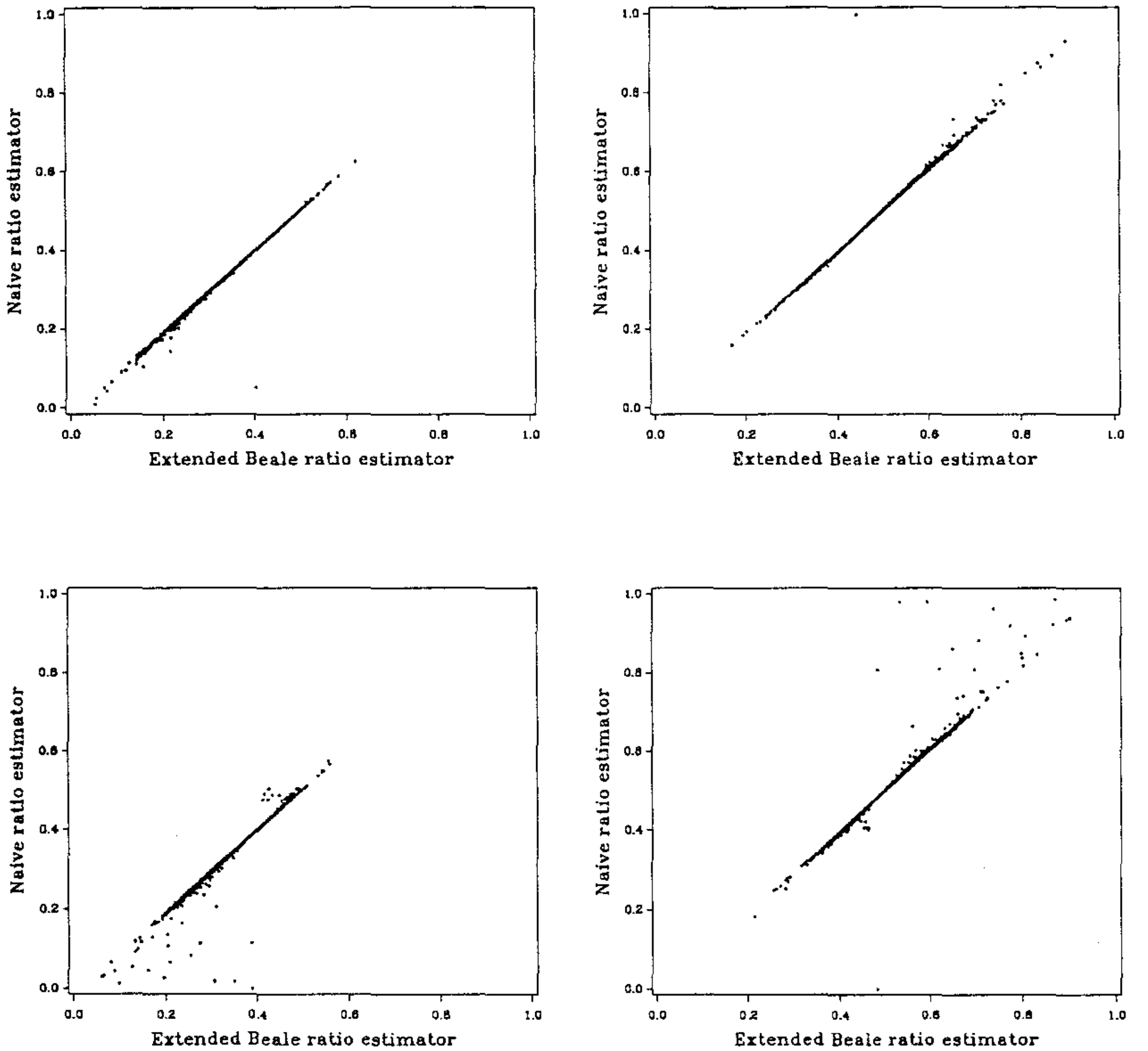

Figure 1. Naive $\hat{R}_{ij}$ vs. Beale $\tilde{R}_{ij}$ ratio estimators for (a) females 20 - 59, total fat, (b) females 20-59, carbohydrates, (c) persons 0 - 19, total fat, and (d) persons 0 - 19, carbohydrates.

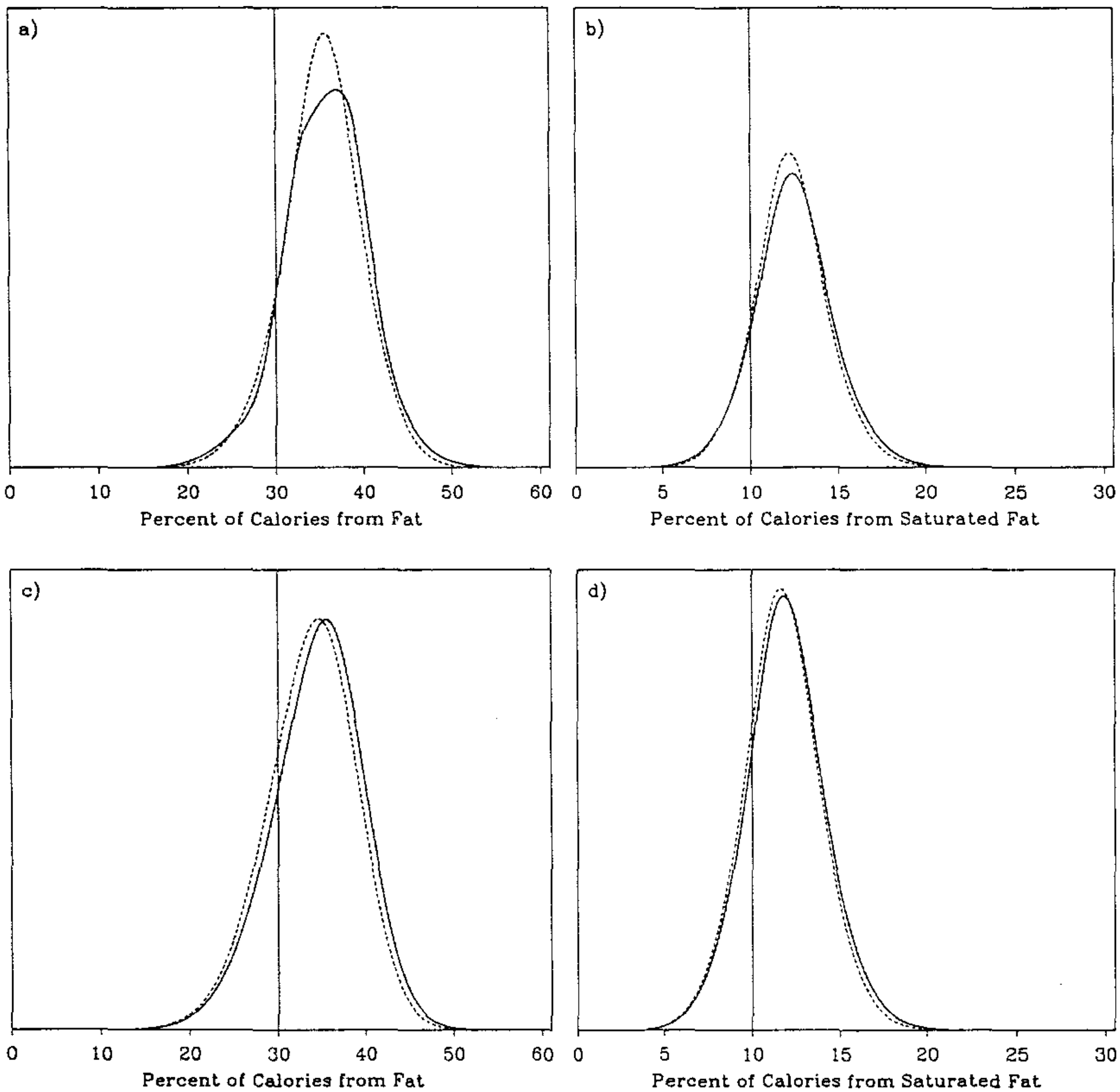

Figure 2. Estimated usual intake distributions from the naive ratio estimates $\hat{R}$ (dotted line) and the Beale ratio estimates $\tilde{R}$ (solid line). (a) Males 20-59, total fat, (b) Males 20-59, saturated fat, (c) Females 20-59, total fat, and (d) Females 20-59, saturated fat