4798.3

MATHEMATICA
Policy Research, Inc.

Contract No.: 53-3 198-7-3 1
MPR Reference No.: 7890-013

THE ROBUSTNESS OF ESTIMATES OF THE DISTRIBUTION OF USUAL DIETARY INTAKE TO ALTERNATIVE SAMPLE SIZES

December 16, 1992

Authors:

Sheena McConnell

Submitted to:

U.S. Department of Agriculture
Office of Analysis and Evaluation
Food and Nutrition Service
3101 Park Center Drive
Alexandria, VA 22302

Project Officer: Christine Kissmer

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Avenue, SW.
Suite 550
Washington, DC. 20024

Project Director: Thomas Fraker

ACKNOWLEDGMENTS

The author would like to thank Gary Bickel, Steven Carlson, and Pat McKinney of the Food and Nutrition Services, Peter Basiotis of the Human Nutrition Information Service, and Thomas Fraker and Met-rile Sing of Mathematica Policy Research, Inc., for valuable comments on earlier drafts of the report. Ann Huff and Esther Miller provided expert research assistance. Tom Good edited the report and Sharon Clark prepared the manuscript.

CONTENTS

CONTENTS (continued)

TABLES

FIGURES

EXECUTIVE SUMMARY

Policy-makers use data from food consumption surveys to assess the need for programs to improve the diets of certain population groups and to assess the effectiveness of existing food assistance programs at improving the dietary intake of participants. These assessments typically entail estimating the proportion of the population who are "at risk" of diet-related health problems. The variable of interest in these studies is not a person's dietary intake on any one day, but his or her "usual" or average dietary intake taken over a period long enough to affect the person's health. But, because most food consumption surveys collect only a small number of days of dietary intake data from each person, usual dietary intake can rarely be measured directly. Instead, researchers typically use the average of a person's observed daily dietary intake over several days as a proxy for his or her usual dietary intake.

Although the average observed dietary intake is generally an unbiased estimator of usual dietary intake, the variation across people in average daily dietary intake is larger than the variation in usual dietary intake. Hence, the distribution of average dietary intake has a lower peak and fatter tails than does the distribution of usual dietary intake. This overestimate of the variation in usual dietary intake has important implications for deriving estimates of nutritionally at-risk populations on the basis of the Recommended Dietary Allowances (RDA). If the RDA is to the left of the median of the distribution of average dietary intake, the proportion of the population at risk of nutrient deficiency will be overestimated; if the RDA is to the right of the median, the proportion of the population at risk of nutrient deficiency will be underestimated.

The Subcommittee on Criteria for Dietary Evaluation of the National Research Council (NRC), proposed a statistical procedure for correcting the bias in the variance of the distribution of average dietary intake. The procedure entails estimating an "adjustment factor" that pulls each point in the distribution of average dietary intake towards the mean. The variance of this adjusted distribution of average dietary intake is an unbiased estimator of the variance of usual dietary intake. The NRC adjustment procedure can be applied with as few as two days of dietary intake data from a subset of the persons in a sample.

The purpose of this report is to assess whether "good" estimates of the distribution of usual dietary intake can be obtained with only two or three days of dietary intake data from all persons in a sample, or alternatively with four days of dietary intake data from some persons in the sample and one day of dietary intake data' **from** the remainder of the sample. Significant cost savings could be realized if fewer days of dietary intake data were collected from at least some persons in a food consumption survey.

We estimate distributions of the usual dietary intake of seven dietary components--vitamin C, calcium, iron, protein, food energy, cholesterol, and saturated fat--with samples of four days of dietary intake data from 600 women and 600 children taken from the 1985 Continuing Survey of Food Intakes by Individuals. We reestimate the distributions by using different subsamples of the data, consisting of all 600 people in each sample but with fewer days of dietary intake data from at least some of the persons. Estimates of the distributions made using the NRC adjustment procedure and the full samples of four days of dietary intake data from all 600 women and children are assumed to be our "best"-estimates and are used as a benchmark by which to judge the precision of the estimates made using the subsamples.

Our estimates confirm the importance of using the NRC adjustment procedure. Eight of 14 distributions estimated without the NRC adjustment procedure but with the full sample of dietary intake data yield estimates of the proportion of the sample at nutritional risk that differ from our best estimates by more than 2.5 percentage points. Distributions estimated with the NRC adjustment procedure but with a subsample that is half the size of the full sample yield estimates closer to our best estimates than those derived from the full sample but without the NRC adjustment procedure.

Our findings suggest that fairly good estimates of the distribution of usual dietary intake can be derived with only three days of dietary intake data from everyone in the sample if the NRC adjustment procedure is used. All but one of the estimates of the proportion of the sample at nutritional risk that are derived from a subsample of three days of dietary intake data from all 600 persons differed by less than 2.5 percentage points from our best estimates. We obtained less accurate estimates by using only two days of dietary intake data from all 600 people. Four estimates of the proportion of the sample at nutritional risk that were derived from subsamples that contain two days of dietary intake data from everyone in the sample differed from our best estimates by more than 2.5 percentage points.

We also found that fairly good estimates can be derived by using a sample that contains four days of dietary intake data from only 400 persons in the sample and one day of dietary intake data from the remaining 200 persons in the sample if the NRC adjustment procedure is used. Only two estimates of the proportion of the sample at nutritional risk that are derived from these subsamples differed from our best estimates by more than 2.5 percentage points.

Our worst estimates using the NRC adjustment procedure are those that are derived from subsamples that contain four days of dietary intake data from 200 persons in the sample and one day of dietary intake data from the remainder of the sample. Half of the estimates of the proportion of the sample at nutritional risk that are derived from subsamples that contain four days of dietary intake data from only 200 persons in the sample differed from our best estimates by more than 2.5 percentage points.

Our results suggest that food consumption surveys with limited budgets should collect two or three days of dietary intake data from each person in a sample, rather than a larger number of days of dietary intake data only from a subset of the persons in the sample. Estimates derived from two days of dietary intake data from all 600 persons were generally better than estimates derived from four days of dietary intake data from 200 persons and one day of dietary intake data from 400 persons--even though both samples contain the same number of observations. Estimates derived from three days of dietary intake data from all 600 persons were similar to estimates derived from four days of dietary intake data from 400 persons and one day of dietary intake data from 200 persons.

The size of the sample required to make good estimates of the distribution of usual dietary intake varies by dietary component. Estimates of the distributions of calcium, iron, protein, and cholesterol intake by women and protein, energy, and saturated fat intake by children were fairly robust to changes in the size of the sample. However, estimates of the distribution of vitamin C intake by women and the distribution of vitamin C and cholesterol intake by children were sensitive to the size of the sample.

We found that estimates of the mean usual dietary intake increased when the sample contained a higher proportion of daily dietary intake reported in the first interview rather than subsequent interviews. A higher estimate of the mean of the distribution of usual dietary intake affects the estimate of the proportion of the sample at risk of either an inadequate or an excess intake of a dietary component. Hence, the difference in the estimates of the proportion of the sample at nutritional risk across samples is due to differences in estimated means in addition to differences in the NRC adjustment factor used.

Two limitations of this study should be kept in mind. First, the study estimates distributions based on subsamples of data only for seven dietary components in two separate subsamples. Hence, the results may not be generalizable to estimates of the distributions of the usual intake of other dietary components by other groups of individuals. Second, we compare our estimates made from the subsamples of the data with ***estimates*** of the distributions derived by applying the NRC adjustment procedure to the full sample rather than the ***true*** distributions. But, if the assumptions underlying the NRC adjustment procedure are not met, the full sample estimates may not be the ***best*** estimates of the distributions of usual dietary intake. To the extent that the full sample estimates differ from the true distributions, the usefulness of the comparisons made in this study to food consumption surveys is limited.

1. INTRODUCTION

For many years the US. Department of Agriculture (USDA) has supported the collection of survey data on food consumption. Policy-makers use data on dietary intake by individuals to assess the need for programs to improve the nutritional quality of the diets of certain population groups and to assess the effectiveness of existing food assistance programs at improving the dietary intake of participants. These assessments typically entail estimating the proportion of the population who are "at risk" of diet-related health problems.

A. STATEMENT OF THE PROBLEM

In general, a person's health is determined not by a deficiency or an excess of a dietary component on any particular day, but by his or her average level of dietary intake over a longer period of time.[1] Thus, for most studies, the variable of interest is not dietary intake on any one day, but the "usual" or average dietary intake over a period long enough to affect a person's health.

In most food consumption surveys, data on the dietary intake of individuals are collected only for a few days. But, due to day-to-day variation in food consumption, a person's dietary intake on any one day may differ from his or her usual dietary intake. Thus, usual dietary intake is rarely measured directly. Instead, researchers typically use the ***average*** of a person's observed daily dietary intake over several days as a proxy for his or her usual dietary intake.

The inability to observe usual dietary intake is particularly a problem with studies that focus on the distribution of usual dietary intake. Although, in most cases, the average observed daily dietary intake is an unbiased estimator of usual dietary intake, the variation across people in average daily dietary intake is larger than the variation in usual dietary intake. Hence, the variance of average

‘An exception is the intake of toxic substances.

dietary intake is a biased estimate of the variance of usual dietary intake. Consequently, the distribution of average dietary intake is flatter than the distribution of usual dietary intake.

This overestimate of the variation in usual dietary intake has important implications for deriving estimates of nutritionally at-risk populations on the basis of the Recommended Dietary Allowances (RDA) or other cutoff points. The distribution of average dietary intake has a lower peak and fatter tails than does the distribution of usual dietary intake. Hence, if the cutoff is to the left of the median of the distribution of average dietary intake, the proportion of the population at risk of nutrient deficiency will be overestimated, if the cutoff is to the right of the median, the proportion of the population at risk of nutrient deficiency will be underestimated. Thus, using the distribution of average dietary intake could lead to serious over- or underestimates of the size of the nutritionally at-risk population.

One solution to the problem of estimating the distribution of usual dietary intake is to collect a large number of days of dietary intake data from survey respondents in order to reduce the bias in the variation of the distribution to an "acceptable" level. However, because studies suggest that at least six days of data are required for many dietary components (Basiotis et al., 1987), this strategy can be prohibitively expensive: not only does it require more interviews with each. person in the sample, but, because response rates fall in later waves of interviewing, it also requires a larger initial sample of persons.

The Subcommittee on Criteria for Dietary Evaluation of the National Research Council (NRC) proposed a statistical procedure for correcting the bias in the variance of the distribution of average dietary intake. The procedure entails estimating an "adjustment factor" that pulls each point in the distribution of the average dietary intake towards the mean. The variance of this adjusted distribution of average dietary intake is an unbiased estimator of the variance of usual dietary intake. An advantage of this procedure is that it can be applied with only two days of dietary intake data from either all of the respondents in a sample or from a subset of the respondents (National Research Council, 1986).

B. THE PURPOSE OF THIS REPORT

A possible implication of the NRC adjustment procedure is that "good" estimates of the distribution of usual dietary intake can be derived with only a few days of dietary intake data from each person in the sample, or with a larger number of days of dietary intake data from a subset of the sample and just a single day of data from the remaining persons in the sample. Using the NRC adjustment procedure would significantly reduce the cost of food consumption surveys if good estimates of the distribution of usual dietary intake could be derived with fewer days of intake data from at least some persons.

The purpose of this report is to assess whether the NRC adjustment procedure can be used to generate good estimates of the distribution of usual dietary intake with only two or three days of dietary intake data from each person, or alternatively with a larger number of days of dietary intake data from a subsample of those persons and one day of data from the remaining persons in the sample. It also addresses whether a food consumption survey with a limited budget for interviews would produce better estimates of the distribution of usual dietary intake if everyone in the sample were interviewed at least twice or if a small sample of people were interviewed four times and the remainder only once.

The report addresses three questions:

- How many days of dietary intake data are required from each person in the sample to obtain good estimates of the distribution of usual dietary intake with the NRC adjustment procedure?

- From what proportion of the sample is more than one day of dietary intake data necessary to obtain good estimates of the distribution of usual dietary intake with the NRC adjustment procedure?

- Can better estimates of the distribution of usual dietary intake be obtained with the ***same*** number of interviews if (1) a few days of data were collected from the entire sample or (2) a larger number of days of data were collected from a small proportion of the sample and one day of data from the remainder of the sample?

The answers to these questions have implications for the design of food consumption surveys.

Our estimates of the distribution of usual dietary intake are derived from a sample of 600 women and a sample of 600 children from the 1985 Continuing Survey of Food Intakes by Individuals (CSFII). The samples are restricted to women and children for whom at least four days of dietary intake data are available. We use the NRC adjustment procedure to estimate the distribution of usual dietary intake under different assumptions about (1) the size of the subsample for whom multiple days of intake data are available, and (2) the number of days of intake data available for each person in the subsample.

To judge the quality of each estimate of the distribution of usual dietary intake, we would ideally compare each of our estimates with the true distribution of usual dietary intake. Because the true distribution is unknown, we use the distribution of dietary intake estimated with the NRC adjustment procedure and all four days of dietary intake data for each person in the sample as our best estimate of the true distribution. We judge the quality of our estimates by comparing the shape of each estimated distribution of usual dietary intake with the shape of our best estimate of the true distribution. The comparison is based on three statistics: (1) the mean of the distribution, (2) the standard deviation of the distribution, and (3) the proportion of the sample whose dietary intake is below the RDA or above the maximum dietary guideline. To check the robustness of our results, we estimate distributions for seven dietary components and for both the sample of women and the sample of children.

We stress that the purpose of the report is to discuss the methodology of estimating distributions of usual intake and not to provide reliable estimates of the proportion of the population at risk of diet-related health problems.

C. OUTLINE OF REPORT

Chapter II of the report discusses the problems associated with estimating the distribution of usual dietary intake with only a few days of dietary intake data per person. Chapter III describes the adjustment procedure recommended by the NRC and the assumptions underlying the procedure. Chapter IV describes the data used in the study. Chapter V presents our best estimates of the distribution of usual dietary intake based on four days of intake data from each person in the sample. In Chapter VI, we compare the estimates of the distribution of usual dietary intake based on subsamples of the data with our best estimates. Chapter VII summarizes our results and discusses their implications for the design of food consumption surveys.

II. PROBLEMS ASSOCIATED WITH ESTIMATING USUAL DIETARY INTAKE

The USDA uses dietary intake data to estimate the proportion of the population who are at risk of health problems from an inadequate or an excess consumption of a dietary component. For this purpose, the variable of interest is a person's usual dietary intake or average dietary intake over a period long enough to affect the person's health.[1] Due to the expense of collecting dietary intake data, most food consumption surveys collect only a few days of dietary intake from each person. The purpose of this chapter is to describe the problems associated with estimating distributions of usual dietary intake with only a few days of dietary intake data from each person in the sample.

A. INTERINDMDUAL AND INTRAINDMDUAL VARIATION

It is helpful to think of a person's observed daily intake of a dietary component as the sum of the person's usual dietary intake and a random error. Thus, the observed daily dietary intake of a person i at time t, y_{it}, can be written as:

II.1 $$y_{it} = x_i + e_{it},$$

where x_i is the usual dietary intake of person i, and e_{it} is a random error whose expected value is zero. The error captures the difference between a person's observed daily dietary intake and his or her usual dietary intake. Assuming that, in the population, a person's usual dietary intake is not correlated with the error, the variation in observed daily dietary intake can then be decomposed into the variation in usual intake and the variation in the error:

11.2 $$V(y_{it}) = V(x_i) + V(e_{it})$$

where V(.) denotes a variance. The variation in usual dietary intake among persons, $V(x_i)$, is known as ***inter-individual*** variation. This is the variation of interest in studies that assess the size of the

[1] The length of time depends on the ability of the body to store nutrients, and varies according to the dietary component (Johnson et al., 1988).

population who are at risk of diet-related health problems. The variation in an individual's observed daily dietary intake around the usual dietary intake, $V(e_{it})$, is known as *intraindividual* variation. Intraindividual variation captures both natural day-to-day variation in diet and variation associated with errors in how dietary intake is measured such as the misreporting of quantities of food consumed.

B. THE IMPLICATIONS OF THE EXISTENCE OF INTRAINDIVIDUAL VARIATION

Estimating the proportion of the population at nutritional risk requires a knowledge of the distribution of usual dietary intake. However, with survey data that include only a few days of dietary intake data from each person, usual dietary intake cannot be observed. Typically, researchers use the average of observed daily dietary intake as a proxy for usual dietary intake.

Unfortunately, the shape of the distribution of average dietary intake differs from the shape of the distribution of usual dietary intake.[2] While the mean of the distribution of average dietary intake is an unbiased estimator of the mean of usual dietary intake, the variance of average dietary intake is larger than the variance of usual dietary intake.

The difference between the distribution of average observed dietary intake and the distribution of usual dietary intake can be understood by taking the average over all days of data of both sides of equation (11.1) yielding:

II.3
$$y_i = x_i + \frac{\sum_t e_{it}}{T},$$

where y_i denotes the average observed dietary intake,

[2]The fundamental problem is that usual dietary intake can be measured only with error. The problems associated with estimating the variance of variables measured with error are well known (see, for example, Maddala, 1979).

$$y_i = \frac{\sum_t y_{it}}{T},$$

and T is the number of days of dietary intake observed for each person. Because the expected value of the error is zero, the expected value of the person's average dietary intake, y_i, is equal to the person's usual dietary intake, x_i. Thus, y_i is an unbiased estimator of %--the mean of the distribution of average observed dietary intake does not vary systematically from the mean of the distribution of usual dietary intake.

However, the variance of average observed dietary intake is a biased estimator of the variance of usual dietary intake. Taking the variance of both sides of equation II.3 yields:

II.4 $$V(y_i) = V(x_i) + \frac{V(e_{it})}{T}.$$

The variance of the average observed dietary intake, $V(y_i)$, exceeds the variance of the usual dietary intake, $V(x_i)$, by the variance of the error, $V(e_{it})$, divided by the number of days of data collected from each person, T. Thus, unless no intraindividual variation exists in the observed data--that is, $V(e_{it})$ = O--or unless an infinite number of days of data are collected from each person, the variance of average observed dietary intake is an overestimate of the variance of usual dietary intake. As the number of days of dietary intake data collected from each person increases, the size of the bias decreases. However, it is important to note, that an increase in the size of the sample without an increase in the number of days of data collected from each person, will not reduce the bias in the variance.

The overestimation of the variance of usual dietary intake due to intraindividual variation has important implications for estimating the proportion of the population at risk of either an inadequate or an excess consumption of a dietary component. Dietary standards such as the RDAs are given in terms of usual intake. The overestimate of the variance of usual dietary intake implies that a greater proportion of the population will seem to fall into the tails of the distribution than is actually the case. Hence, estimates of the proportion of the population that are greater than or less than the dietary standard that are based on the distribution of average dietary intake will be biased.

Figure II.1 illustrates the distribution of average observed dietary intake and the distribution of usual dietary intake for a nutrient. In this example, the RDA cutoff lies to the left of the median of the distribution, and thus an estimate of the proportion of the population with a dietary intake less than the RDA that is derived from the distribution of average observed dietary intake will overestimate the true proportion of the population at risk of nutrient deficiency. If the RDA cutoff lies to the right of the median, an estimate of the proportion of the population with a dietary intake below the RDA will underestimate the proportion of the at risk population. Conversely, the distribution of average observed dietary intake will underestimate the prevalence ofoverconsumption if the recommended maximum dietary intake lies to the left of the median, and will overestimate the prevalence of overconsumption if the recommended maximum dietary intake lies to the right of the median.

The existence of intraindividual variation will also bias downwards measures of correlation between dietary intake and other variables (Liu et al., 1978). This downward bias may lead to the false conclusion that important relationships between dietary intake and other outcomes, such as disease or program participation, do not exist.

C. HOW IMPORTANT IS THE BIAS?

The bias in the estimate of the variance of the distribution of usual dietary intake depends on two factors: (1) the size of the intraindividual variation, and (2) the number of days of dietary intake data collected from each person.

Table II.1 presents estimates of the ratio of intraindividual variation to interindividual variation that are found in the literature. Although the degree of intraindividual variation varies for different dietary components, and in different population groups, intraindividual variation is nearly always larger than interindividual variation. The ratio is lowest for calcium and carbohydrate and highest

FIGURE II.1

COMPARISON OF A **DISTRIBUTION** OF AVERAGE **DIETARY** INTAKE AND **A DISTRIBUTION** OF USUAL **DIETARY** INTAKE

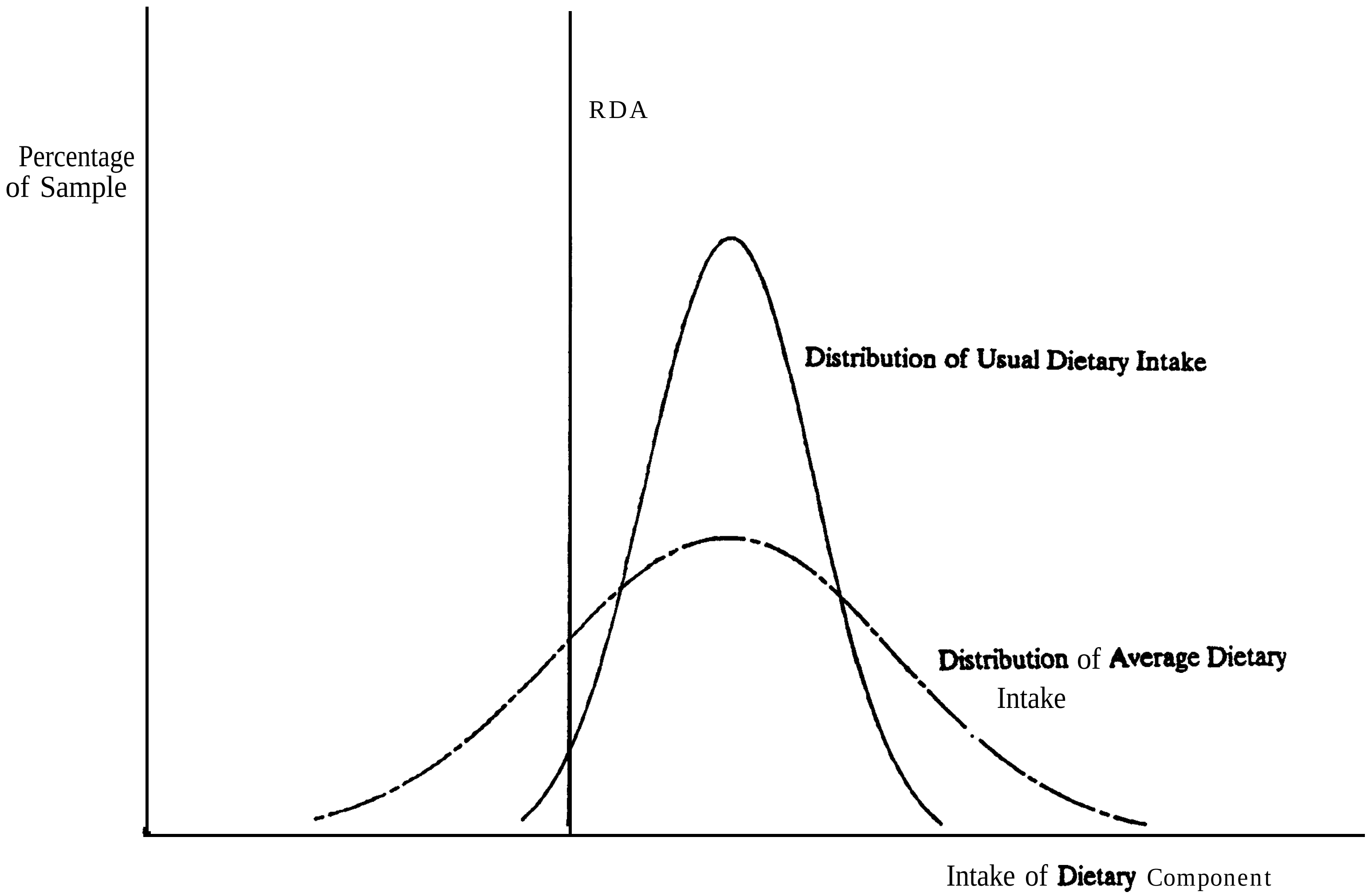

TABLE II.1

OBSERVED RATIOS OF INTRAINDIVIDUAL VARIATION
TO INTERINDIVIDUAL VARIATION

Dietary Component	Range of Estimates
Vitamin C	1.4 - 3.5
Calcium	0.9 - 2.2
Iron	1.3 - 2.7
Protein	1.2 - 2.9
Energy	0.8 - 2.2
Cholesterol	1.6 - 5.6
Percentage of Energy Derived from Saturated Fat	2.1 - 3.9
Carbohydrate	0.6 - 2.1
Vitamin A	1.6 - 24.3
Thiamin	0.9 - 4.4
Total Fat	0.9 - 1.6

SOURCE: Battese et al. (1988), Fraker, Long, and Post (1990), Liu et al. (1978), National Research Council (1986), and Ritenbaugh et al. (1988).

for vitamin A and cholesterol. Dietary components that are consumed regularly, such as calcium, tend to have a low ratio of intraindividual to interindividual variation, while those that are consumed irregularly, such as vitamin A, have a high ratio of intraindividual to interindividual variation (Sempos et al., 1985). Sempos et al. also found that for many dietary components the ratio of intraindividual to interindividual variation falls when dietary intake is defined to include vitamin and mineral supplements. This broader definition of dietary intake is generally not used in USDA food consumption surveys.

Basiotis et al., (1987) Liu et al. (1978), the National Research Council (1986), Sempos et al. (1985), and Ritenbaugh et al. (1988) have estimated the number of days of data required from each person to obtain estimates with any given degree of accuracy of the distribution of usual dietary intake using the conventional approach. The number of days required varies among dietary components and data sets. However, most studies have found that at least several days of data are necessary for ensuring a reasonable degree of accuracy for nutrients that exhibit a low ratio of intraindividual to interindividual variation, such as food energy, and weeks of data are required for nutrients that exhibit a high ratio of intraindividual to interindividual variations, such as vitamin A and thiamin.

Basiotis et al. used a sample containing 365 days of intake data from 29 persons to calculate the number of days of intake data required from each person to estimate, using the conventional approach, the average usual intake of each dietary component within 10 percent of the "true" average 95 percent of the time. The estimate made using all 365 days of intake data was assumed to be the "true" average. The study found that 3 days of intake data are required to estimate the average usual intake of food energy among both men and women, but 39 days of intake data are required to estimate the average usual intake of vitamin A among men and 44 days are required to estimate the average usual intake of vitamin A among women.

Collecting many days of intake data from each person in the sample can be prohibitively expensive for two reasons. First, collecting additional days of intake data usually requires recontacting the person either by telephone or in-person. (Alternatively, extra days of intake data can be collected by asking the person to record their food consumption in a diary. However, this adds additional burden to the respondent, increases the reporting error, and requires that food consumption is recorded on consecutive days). Second, there is evidence that respondents are much less likely to respond to later requests for dietary intake data. Hence, to allow for attrition from the sample, if more days of intake data are required a larger initial sample of persons must be interviewed to collect a full set of dietary intake data from a target number of respondents.

III. THE NRC ADJUSTMENT PROCEDURE

Because food consumption surveys rarely collect more than a few days of data on dietary intake from each person, the distribution of usual dietary intake must be estimated from the average value of observed dietary intake. But, as discussed in the previous chapter, the presence of intraindividual variation means that the variation in average observed dietary intake exceeds the variation in usual dietary intake. Not accounting for this difference in variation can cause serious errors in estimates of the proportion of the population at nutritional risk. In response, the NRC proposed a procedure that adjusts the distribution of average observed dietary intake to take into account intraindividual variation (National Research Council, 1986). This chapter describes the adjustment procedure proposed by the NRC. Appendix A provides a more detailed and technical discussion of the procedure. Appendix B discusses two alternative procedures for estimating the distribution of usual dietary intake.

A. ESTIMATING THE DISTRIBUTION OF USUAL DIETARY INTAKE

The NRC adjustment procedure is based on the assumption that the distribution of average observed daily dietary intake has the same mean and a similar shape as does the distribution of usual dietary intake, but that the variance, or spread, of the distribution of average observed dietary intake is larger than the variance of the distribution of usual dietary intake. The NRC procedure estimates the distribution of usual dietary intake by moving each point in the distribution of average observed dietary intake toward the mean of the distribution. Thus, it produces a distribution whose spread is smaller than the spread of the distribution of average observed dietary intake.

The distance that each point is moved towards the mean of the distribution is determined by an adjustment factor. The adjustment factor depends on the relative size of the intraindividual and interindividual variation and the number of days of dietary intake observed for each person in the sample. The larger the relative size of the intraindividual variation, the larger the proportion of the variance of average observed dietary intake that is explained by day-to-day variation in dietary intake rather than by the variation among individuals, and thus the farther the adjustment procedure pulls each point in the distribution of average observed dietary intake toward the mean to estimate the distribution of usual dietary intake. Similarly, the greater the number of days of dietary intake observed for each person, the closer the average observed dietary intake is to usual dietary intake, and the smaller the degree of adjustment necessary to estimate the distribution of usual dietary intake.

We denote the average value of the observed dietary intake by each person as y_i, and the overall average of the observed dietary intake by y. A point on the estimated distribution of usual dietary intake is denoted by: $\hat{x}_i$ The formula proposed by the NRC to estimate the distribution of usual dietary intake, using this notation is:

III.1 $$\hat{x}_i = y + (y_i - y) * \textit{Adjustment Factor.}$$

The adjustment factor is given by the formula:

III.2 $$\textit{Adjustment Factor} = \sqrt{\frac{V(x_i)}{V(x_i) + \frac{V(e_{it})}{T}}},$$

where $V(x_i)$ measures interindividual variation, $V(e_{it})$ measures intraindividual variation, and T measures the number of days of dietary intake collected from each person. The adjustment factor is sometimes known as the "attenuation factor."

The mean of the estimated distribution of usual dietary intake given in equation (111.1) is y, which is the mean of the distribution of average observed dietary intake.[1] However, except at the mean, $\hat{x}_i$ is not equal to the average observed dietary intake, y_i. Because the adjustment factor can

[1]In equation (111.1) note that $\hat{x}_i$ = y when $y_i = y$.

never be greater than 1, each point in the estimated distribution of usual dietary intake, $\hat{x}_i$, will be closer to the mean than will the corresponding point in the distribution of average observed dietary intake, y_i. Hence the tails of the estimated distribution of usual dietary intake will be thinner than the tails of the distribution of average observed dietary intake. For persons with dietary intake below the average, the estimate of usual dietary intake using the adjustment factor will be greater than the average observed dietary intake. Conversely, for persons with dietary intake above the average, the estimate of usual dietary intake will be lower than the average observed dietary intake.

If intraindividual variation did not exist--that is, $V(e_{it}) = 0$ in the denominator of equation (III.2)--the adjustment factor would equal 1, and the estimated distribution of usual dietary intake would be the distribution of average observed dietary intake. Moreover, as the number of days of dietary intake collected from each person, T, increases, the adjustment factor increases towards 1, and the estimated distribution of usual dietary intake comes closer to the distribution of average observed dietary intake.

B. ESTIMATING INTRAINDMDUAL AND INTERINDIVIDUAL VARIATION

The adjustment factor is a function of intraindividual and interindividual variation--both of which are unobservable. However, they can be estimated from an analysis of variance (ANOVA) regression in which the dependent variable is observed *daily* dietary intake of individual i at time t, y_{it}, the independent variables are dummy variables, one for each person in the sample, which equal 1 if the observed dietary intake is by the i-th person and 0 otherwise. ANOVA computer packages calculate the variance of the observed daily dietary intake (MSMODEL) and the variance of the observed daily dietary intake not explained by the dummy variables (MSERROR). The MSMODEL and MSERROR can be used to estimate intraindividual variation, $V(e_{it})$, and interindividual variation, $V(x_i)$. The formulas for the estimates are as follows:

$$\textit{Estimated intraindividual variation} = MSERROR$$

and

$$\textit{Estimated interindividual variation} = \frac{MSMODEL - MSERROR}{T}$$

Substituting these estimates into the formula for the adjustment factor (equation (III.2) yields:

$$\textit{Estimated Adjustment Factor} = \sqrt{\frac{MSMODEL - MSERROR}{MSMODEL}}$$

C. THE ASSUMPTIONS UNDERLYING THE NRC ADJUSTMENT PROCEDURE

The NRC procedure is based on two assumptions.

Assumption 1

The observed daily dietary intake, y_{it}, is the sum of usual dietary intake and an error,

$$y_{it} = x_i + e_{it},$$

and the error has the following properties:

- ***The Mean Value of the Error is Zero.*** This assumption is not valid if dietary intake is systematically misreported--for example, if (1) persons forget to report some food consumption (for example, food consumed as snacks), (2) persons do not wish to admit to consuming less socially acceptable food, such as alcohol, or (3) persons are interviewed only on days in which food consumption is atypical, (for example, Thanksgiving). If this assumption is not valid, then the mean of the distribution of average observed dietary intake will be a biased estimate of the mean of the distribution of usual dietary intake.

- The ***Error Is Uncorrelated with the Usual Dietary Intake.*** This assumption is not valid if the difference between the observed daily dietary intake and the usual dietary intake varies systematically with the magnitude of the usual dietary intake--for example, if persons who consume more than the average amount of food underreport their consumption, and those who consume less than the average amount of food overreport their consumption.

- The ***Error Is Not Correlated with the Error on Previous or Subsequent Days.*** This assumption is unlikely to be valid in data collected in surveys that observe dietary intake on consecutive days, such as the National Food Consumption Survey (NFCS). Dietary intake may be positively correlated across consecutive days because (1) the days are in the same season, (2) the food remaining from one meal may be served the next day, (3) the food consumed may have been purchased at the same time and place, and (4) an illness or a special diet may persist over several days. Dietary intake may also be negatively correlated across consecutive days if, for example, a day of heavy eating is followed by a day of light eating. However, this assumption is likely to be valid in data collected by such surveys as the 1985 CSFII, in which the days on which the survey observes dietary intake are separated by about two months.

- ***The Variance of the Error Is Constant.*** This assumption requires that the magnitude of intraindividual variation be the same for all individuals. Estimates of the ratios of intraindividual to interindividual variation suggest that they do vary by sex (National Research Council, 1986). Ritenbaugh et al. (1988) found evidence that the variance ratio of vitamin A intake by low-income Hispanics was much higher than the variance ratio for vitamin A intake by other ethnic groups. Ritenbaugh et al. and Johnson et al. (1988) also found that the variance ratio of the intake of vitamin A, and to a lesser extent of other nutrients, depends on the average intake. However, with these exceptions, most studies have shown that variance ratios are remarkably similar across substrata of the population.

Assumption 2

The distributions of both average observed dietary intake, y_i, and usual dietary intake, x_i, are normal. Moreover, each point in the distribution of average dietary intake is the same number of standard deviations away from the mean as is the corresponding point in the distribution of usual dietary intake.

While the NRC adjustment procedure is based on an assumption of normality, it is designed to be robust to small deviations from normality in the distribution of average dietary intake.* The NRC adjustment procedure produces an estimate of the distribution of usual dietary intake by adjusting each point in the distribution of average dietary intake toward the mean. Hence, deviations from normality in the distribution of average dietary intake will be replicated in the estimated distribution of usual dietary intake. For example, if the distribution of average dietary

*Given the assumption of normality, the complete distribution of usual dietary intake could be constructed using the estimate of the mean and the standard deviation and a table of the normal distribution. But, to allow for deviations from normality in the distribution of usual dietary intake, the NRC adjustment procedure does not construct the estimated distribution of usual dietary intake in this manner.

intake has a long right-hand tail, the distribution of usual dietary intake estimated with the NRC adjustment procedure will also have a long right-hand tail.

However, the NRC adjustment procedure is only appropriate if the distribution of usual dietary intake differs from normality in the same way as the distribution of average dietary intake. This is unlikely to be the case if the distribution of average dietary intake deviates substantially from normality. If, for example, the distribution of average dietary intake is skewed or exhibits kurtosis (fatter tails and a higher peak than the normal), the distribution of usual dietary intake is likely to differ from the distribution of average dietary intake in its degree of skewness and kurtosis. Unfortunately, the National Research Council does not give any guidance on the degree of deviation of the distribution of average dietary intake from normality that is acceptable for appropriate use of the NRC adjustment procedure.

In practice, the distributions of average intake of nearly all dietary components deviate from normality to some degree (Battese et al., 1988). Some dietary components, such as vitamin A, have average intake distributions that are highly skewed to the right (Ritenbaugh et al., 1988). The consensus among researchers is that it is inappropriate to use the NRC adjustment procedure to estimate distributions of usual intake when the average intake distribution is highly nonnormal, as in the case of vitamin A. There is less consensus among researchers as to the appropriateness of using the NRC adjustment procedure when the average intake distribution deviates less markedly from the normal--Battese et al. argue that procedures based on distributional assumptions other than normality yield better estimates of the distributions of usual intake of a number of dietary components (see Appendix B).

IV. A DESCRIPTION OF THE DATA

This chapter describes the sample of dietary intake data used in this study. Section A describes the CSFII survey, Section B describes the sample of women and the sample of children taken from the CSFII, and Section C discusses the dietary components selected for the study.

A. THE 1985 CSFII

The 1985 CSFII was the first of two nationwide longitudinal surveys of the dietary intake of U.S. households which contain at least one adult woman. The survey includes three samples of individuals: (1) the "core" sample, which includes women age 19 to 50 and their children age 1 to 5, (2) the "low-income" sample, which consists of a comparable group of women and children whose household income does not exceed 130 percent of the poverty guidelines, and (3) men age 19 to 50 who reside in households that include at least one woman age 19 to 50. Six interviews were arranged for each of the women at about two-month intervals.[1] The first interview was always conducted in-person; subsequent interviews were conducted by telephone or, if the respondent could not be contacted by telephone, conducted in-person. The women were asked to recall the type and quantity of food that they and their children had eaten in the previous 24 hours. In addition, they were asked for information about both personal and household characteristics, including age, weight, income, program participation, and pregnancy and lactation status.

One criticism of previous dietary intake surveys was that the observations on each individual were not independent. For example, the NFCS collected dietary intake data for three consecutive days. To counter this criticism, the 1985 CSFII was designed so that the observations on each individual vary by the day of the week and the season of the year.

---

[1]The men were interviewed only once.

B. A DESCRIPTION OF THE SAMPLE

The Food and Nutrition Service (FNS) of USDA contracted with Mathematica Policy Research (MPR) to create an analysis file which contains data on women and their children from both the core and low-income samples of the CSFII.[2] The attrition rate of respondents was fairly high in the later waves of the CSFII--only 45 percent of the respondents who completed the first interview also completed the remaining five interviews. In order to create a data set with no missing values, but to maintain a fairly large number of observations on each individual, only households for which at least four days of data are available for each individual were kept in the MPR analysis file. Over 70 percent of respondents completed at least four days of data. Only four waves of data were kept for each individual in the sample--the data from the first interview and data from three randomly chosen interviews from the remaining completed interviews. For this study, 600 nonpregnant, nonlactating women and 600 children were chosen randomly from the MPR analysis file.[3] Because four observations are available on each person, each data set contains 2,400 observations. About 60 percent of the women and 40 percent of the children were from the core group, and the remaining 40 and 60 percent, respectively, were from the low-income group.

C. DIETARY COMPONENTS

We estimate the distributions of the usual intake of seven dietary components:

- Vitamin C
- Calcium
- Iron
- Protein
- Food energy

---

[2]The construction of the analysis file is described in detail in Fraker and Post (1989).

[3]The MPR analysis file contains 1,947 women, 1,718 who were not pregnant or lactating, and 760 children.

- Cholesterol
- Saturated Fat

We chose these food components because a deficient or excessive consumption of them is believed to have an important impact on a person's health. While a deficiency of vitamin C, calcium, iron, or protein can cause health problems, an excessive consumption of cholesterol and saturated fat is associated with certain diseases. Both a deficient and an excessive intake of food energy can cause health problems.

Because individuals with higher food energy intake can consume larger quantities of saturated fat than can those with lower energy intake without adverse effects on their health, we estimate the distribution of the percentage of total food energy intake which is derived from saturated fat rather than the distribution of saturated fat intake.

Table IV.1 presents descriptive statistics on the distributions of the average intake of each dietary component calculated from the four days of intake data. These are presented together with the RDA or maximum intake guideline for each dietary component. The RDA for vitamin C, calcium, iron, and protein is the quantity of the nutrient required to meet the needs of nearly all healthy persons and thus exceeds the requirements of most individuals. Because health risks are associated with both a deficient and an excessive intake of food energy, the RDA for food energy is set at the quantity required to meet the needs of the average individual. RDA for cholesterol and saturated fat have not yet been established; instead, the table shows guidelines for the maximum intake of these food components.

The mean intake of dietary components in these samples are similar to those found in larger samples from the CSFII. Pao et al. (1989) report that the mean intake of food energy calculated from four days of data from each individual is 1,528 kcal for women in the core sample, 1,427 kcal for women in the low-income sample, 1,426 kcal for children in the core sample, and 1,332 kcal for children in the low-income sample. These figures compare with an average energy intake of 1,473

TABLE IV.1

DISTRIBUTIONS OF FOUR-DAY AVERAGE DIETARY INTAKE

Dietary Component	RDA or Dietary Guideline	Mean	Standard Deviation	Skewness	Kurtosis
Sample of 600 Women (2,400 observations)					
Vitamin C (mg)	60	73.6	51.7	1.6	3.6
Calcium (mg)	1,200/800[a]	548.6	286.8	1.1	2.4
Iron (mg)	15	9.7	3.9	1.1	2.7
Protein (g)	46/50[a]	57.7	20.9	0.8	2.4
Energy (kcal)	2,200	1,473.7	512.0	0.6	0.5
Cholesterol (mg)	300	279.5	158.6	1.3	2.8
Percentage of Energy Derived from Saturated Fat (%)	10	12.6	2.8	0.1	-0.1
Sample of 600 Children (2,400 observations)					
Vitamin C (mg)	40/45[b]	80.6	46.6	1.3	2.9
Calcium (mg)	800	761.2	273.6	0.7	1.3
Iron (mg)	10	9.8	4.2	2.6	12.0
Protein (g)	16/24[b]	52.9	15.5	0.9	1.9
Energy (kcal)	1,300/1,800[b]	1,403.4	373.0	0.8	2.0
Cholesterol (mg)	300	250.3	133.5	2.3	14.2
Percentage of Energy Derived from Saturated Fat (%)	10	13.8	2.4	0.1	0.5

SOURCE: Random samples taken from the MPR analysis file that contains four days of intake data from the core and low-income samples of the 1985 CSFII.

NOTE: RDAs and dietary guidelines from ***Recommended Dietary Allowances*** (National Research Council, 1989)

[a]The first figure is the RDA for women age 19 to 24, and the second figure is the RDA for women age 25 to 50.

[b]The first figure is the RDA for children age 1 to 3, and the second figure is the RDA for children age 4 to 8.

kcal for women and 1,403 kcal for children taken over both the core and low-income samples used in this study.

The mean intake of iron and food energy is more than one standard deviation lower than the RDA in the sample of women. In the sample of children, the mean intake of a dietary component is never more than one standard deviation lower than the RDA.[4] However, the mean value of the percentage of energy derived from saturated fat exceeds the intake guideline by more than one standard deviation.

The NRC adjustment procedure is designed to be applied to distributions of average dietary intake that are approximately normally distributed. Two measures of the deviation of each distribution from the normal distribution, skewness and kurtosis, are given in the last two columns of Table IV.1. Skewness measures the degree of symmetry in the distribution. If the distribution of the average intake of a dietary component is a normal distribution (which is symmetric), the measure of skewness will be zero. If the distribution of average intake is skewed to the right (the intake of more than half of the individuals is less than the mean), then the measure of skewness will be positive. Kurtosis is a measure of the shape of the tails of the distribution. If the distribution is normal then the kurtosis measure will be zero; if the distribution has fatter tails and a higher peak than the normal distribution, then the kurtosis measure will be positive.

The distribution of all dietary components considered in this report are skewed to the right. The intake of iron and cholesterol by children exhibit the most highly skewed distributions. The distributions of the intake of protein, food energy, and the percentage of energy derived from saturated fat exhibit very little asymmetry in either of the samples. The distributions of all but one of the dietary components--saturated fat intake by women--have fatter tails than the normal distribution. The distribution of iron and cholesterol intake among children, and vitamin C intake among both women and children, show the greatest deviation from the normal distribution. Other

[4]Except if the mean intake of protein by all children is compared to the RDA for protein intake by children age 4 to 8.

studies have estimated similar distributions using different samples (National Research Council, 1986; Battese et al., 1988; and Fraker, Long, and Post, 1990).

The National Research Council does not provide guidelines on the degree of deviation from normality that is acceptable for use of the NRC adjustment procedure. We choose not to estimate the distribution of vitamin A in this study because in nearly all studies the distribution of vitamin A is found to be highly skewed?

[5] The National Research Council (1986) reports that the measures of skewness and kurtosis are 3.9 and 20.7, respectively, for the distribution of three-day averages of vitamin A calculated from a sample of women from the 1977-1978 NFCS.

V. ESTIMATES OF THE DISTRIBUTIONS OF USUAL DIETARY INTAKE DERIVED FROM THE FULL SAMPLE

This chapter discusses estimates of the distributions of usual dietary intake derived from four days of intake data for each of the 600 persons in the two samples of CSFII participants that we selected for this study. These estimates are our "best" estimates of the distributions and, in the next chapter, they form the basis for judging the quality of the estimates of the distributions derived from subsamples of the data. To estimate the distribution of usual dietary intake, we adjusted the average dietary intake for two sources of variation: (1) variation due to differences in the type and timing of the interviews, and (2) other unobservable intraindividual variation. We discuss each in turn.

A. ADJUSTING FOR THE TYPE AND TIMING OF THE INTERVIEW

Previous studies have found that the reported intake of a dietary component varies systematically with the day of the week on which the person is interviewed, the season in the year on which the person is interviewed, the method used to interview the person, and the number of times that the person has already been interviewed as part of the survey (Battese et al,, 1988, and Ritenbaugh et al., 1988). For example, Battese et al. found that women report a higher consumption of calories on the weekend. These "interview effects" increase intraindividual variation. Thus, they increase the total variation in average dietary intake and exacerbate the problem of estimating the distribution of usual dietary intake.

However, adjusting the data on daily dietary intake to remove the variation due to the type and timing of the interviews is straightforward. To do so, we estimated the magnitude of the interview effects from regression models of the observed daily intake of each dietary component. The explanatory variables in the models include a series of dummy variables that indicate whether (1) the intake was observed on a weekend, (2) the intake was observed in the summer, fall, or winter, (3) the respondent was interviewed in person or by telephone, and (4) the interview was conducted in the first, second, third, fourth, or fifth wave of interviews. We added the residuals from these regressions to the mean intake of the dietary component to form a series of daily dietary intakes adjusted for interview effects.

The interview effects are not large in the sample of dietary intake data used in this study.[1] The R-squared statistic in each of the regression models is extremely low--little variation in observed daily intakes can be attributed to interview effects.[2] However, statistically significant interview effects are found in the reported intake of some dietary components. Women in our sample reported consuming more calories on the weekend, and children in our sample consumed different amounts of protein, energy, and saturated fat according to the season of the year. Roth the samples of women and children reported consuming lower amounts of calcium and vitamin C when interviewed in person rather than by telephone. Reported intake of protein, cholesterol, and saturated fat by children was higher when the interview was in-person. In both samples, the reported intake of energy was higher in the first wave of interviewing than in the later waves. The reported intake of calcium by women was also higher in the first wave of interviewing.

Table V.l presents summary statistics of the distributions of average dietary intake that are adjusted for interview effects but not for unobservable intraindividual variation. A comparison of the means and standard deviations of the unadjusted distributions of average dietary intake in Table IV.1 and the adjusted distributions in Table V.I shows that the means of the distributions are unaltered by the adjustment, but that the standard deviations of the adjusted distributions are smaller than the standard deviations of the unadjusted distributions. However, the impact of the adjustment on the standard deviations is small. The largest changes occur in the standard deviation of the average intake of iron in the sample of women, but the change is less than 3 percent.

---

[1]Appendix C presents the results of these regressions.

[2]Fraker Long, and Post (1990) also found low R-squared statistics in similar regressions in a sample of low-income children from the 1985 CSFII.

TABLE V.1

DISTRIBUTIONS OF FOUR-DAY AVERAGE DIETARY INTAKE: ADJUSTED FOR DAY, SEASON, WAVE, AND METHOD OF CONTACT

Dietary Component	Mean	Standard Deviation	Percentage Below RDA or Above Dietary Guideline
Sample of 600 Women (2,400 observations)			
Vitamin C (mg)	73.6	51.3	49.7
Calcium (mg)	548.6	284.5	84.0
Iron (mg)	9.7	3.8	90.7
Protein (g)	57.7	20.8	0.4
Energy (kcal)	1,473.7	510.1	90.8
Cholesterol (mg)	279.5	157.9	37.0[a]
Percentage of Energy Derived from Saturated Fat (%)	12.6	2.8	83.3[a]
Sample of 600 Children (2,400 observations)			
Vitamin C (mg)	80.6	46.3	22.8
Calcium (mg)	761.2	272.0	61.0
Iron (mg)	9.8	4.2	64.7
Protein (g)	52.9	15.5	0.5
Energy (kcal)	1,403.4	371.6	85.5
Cholesterol (mg)	250.3	131.9	27.2[a]
Percentage of Energy Derived from Saturated Fat (%)	13.8	2.4	94.0[a]

SOURCE: Random samples taken from the MPR analysis file containing four days of intake data from the core and low-income samples of the 1985 CSFII.

[a]Percentage of sample whose average intake is above guideline.

B. ADJUSTING FOR UNOBSERVABLE INTRAINDMDUAL VARIATION

We used the NRC adjustment procedure to adjust the distributions of average dietary intake for unobservable intraindividual variation. We used the daily dietary intake data adjusted for interview effects to estimate intraindividual and interindividual variation and, hence, the NRC adjustment factor. We calculated each person's average dietary intake from the daily dietary intake data adjusted for interview effects. We then used the adjustment factor to adjust the distributions of the average dietary intake for any remaining intraindividual variation.

Tables V.2 and V.3 present the estimates of the ratios of intraindividual to interindividual variation--the variance ratios--and the adjustment factors used to adjust the average dietary intake for intraindividual variation. For each dietary component, in each sample, the intraindividual variation is greater than the interindividual variation--the variance ratio is greater than one. Hence, for each dietary component, the adjustment factor is less than one--the variation in the adjusted average intake is lower than the variation in the unadjusted average intake. Lower variance ratios are associated with higher adjustment factors. The closer the adjustment factor is to one, the smaller the adjustment of the distribution of average dietary intake required to estimate the distribution of usual dietary intake.

In both samples, the variance ratios are lowest for calcium and energy and highest for cholesterol and saturated fat. Most of the estimates of the variance ratios lie in the range of estimates found in previous studies (see Table 11.1). The exceptions are estimates of the variance ratios for saturated fat in both samples and iron and cholesterol in the sample of children; all these estimates are higher than previous estimates.

Tables V.2 and V.3 also present the means and standard deviations of the adjusted distributions and the percentage of the sample whose estimated usual intake is less than the RDAs or more than the maximum dietary guidelines. For comparison, Tables V.2 and V.3 also present the percentage

TABLE V.2

ESTIMATED DISTRIBUTIONS OF USUAL DIETARY INTAKE: ADJUSTED FOR INTRAINDIVIDUAL VARIATION: FULL SAMPLE OF WOMEN

Dietary Component	Variance Ratio	Adjustment Factor	RDA or Dietary Guideline	Mean	Standard Deviation	Percentage Below RDA or Above Guideline (Unadjusted)	Percentage Below RDA or Above Guideline (Adjusted)
Sample of 600 Women (2,400 observations)							
Vitamin C (mg)	3.31	0.74	60	73.6	37.9	49.7	45.3
Calcium (mg)	1.75	0.83	1,200/800[a]	548.6	236.0	84.0	86.5
Iron (mg)	2.46	0.79	15	9.7	3.0	90.7	94.8
Protein (g)	2.84	0.76	46/50[a]	57.7	15.8	0.4	0.3
Energy (kcal)	1.96	0.82	2,200	1,473.7	418.0	90.8	94.3
Cholesterol (mg)	3.87	0.71	300	279.5	112.0	37.0[b]	34.8[b]
Percentage of Energy Derived from Saturated Fat (%)	4.68	0.68	10	12.6	1.9	83.3[b]	92.5[b]

SOURCE: Random samples taken from the MPR analysis file that contains four days Of intake data from the core and low-income samples of the 1985 CSFII.

NOTE: Intakes are also adjusted for day, season, wave, and method of contact.

[a]For the purpose of calculating the proportion of the sample at nutritional risk, the second RDA is used.

[b]Percentage of sample whose average intake is *above* the dietary guideline.

TABLE V.3

ESTIMATED DISTRIBUTIONS OF USUAL DIETARY INTAKE: ADJUSTED FOR INTRAINDIVIDUAL VARIATION
FULL SAMPLE OF CHILDREN

Dietary Component	Variance Ratio	Adjustment Factor	RDA or Dietary Guideline	Mean	Standard Deviation	Percentage Below RDA or Above Guideline *(Unadjusted)*	Percentage Below RDA or Above Guideline *(Adjusted)*
Sample of *600* Children (2,400 observations)							
Vitamin C (mg)	3.12	0.75	40/45[a]	80.6	34.7	22.8	11.0
Calcium (mg)	1.93	0.82	800	761.2	222.9	61.0	62.8
Iron (mg)	3.17	0.75	10	9.8	3.1	64.7	65.2
Protein (g)	2.49	0.78	16/24[a]	52.9	12.0	0.5	0.0
Energy (kcal)	2.02	0.82	1,300/1,800[a]	1,403.4	304.5	85.5	91.2
Cholesterol (mg)	6.71	0.61	300	250.3	80.4	27.2[b]	20.8[b]
Percentage of Energy Derived from Saturated Fat (%)	6.03	0.63	10	13.8	1.5	94.0[b]	99.7[b]

SOURCE: Random samples taken from the MPR analysis file that contains four days of intake data from the core and low-income samples of the 1985 CSFII.

NOTE: Intakes are also adjusted for day, season, wave, and method of contact.

[a]For the purpose of calculating the proportion of the sample at nutritional risk, the higher RDA or guideline is used.

[b]Percentage of sample whose average intake is ***above*** the dietary guideline.

of the sample whose average dietary intake unadjusted by the NRC procedure is less than the RDAS or above the dietary guidelines.

The NRC adjustment procedure does not affect the mean of the distribution of average intake; however, because it removes the intraindividual variation, the standard deviation of the adjusted distribution is smaller. This can be seen by comparison of the means and standard deviations in Tables V.2 and V.3 with those reported in Table V.l.

The change in the standard deviations of the distributions affects the estimates of the proportion of the sample whose average dietary intake are less than the RDAs or above the dietary guidelines. For 8 of the 14 estimated distributions, the change in the estimate of the proportion of the population at risk of either inadequate or excess consumption was greater than 2.5 percentage points. In five cases--saturated fat in the sample of women, andvitamin C, energy, cholesterol, and saturated fat in the sample of children--the change in the estimates of the proportion of the sample either below the RDA or above the dietary guideline was greater than 5 percentage points.

In both samples, the adjusted distributions yield estimates of the proportion of the sample whose intake of calcium, iron, and energy are below the RDA that are higher than those yielded by the unadjusted distribution. On the other hand, in both samples, the adjusted distributions yield lower estimates of the proportion of the samples at risk of vitamin C and protein deficiency. The adjusted distributions yield lower estimates of the proportion of both samples at risk of excess cholesterol intake and a higher estimate of the proportion at risk of excess saturated fat intake.

Whether the adjusted distributions yield larger or smaller estimates of the proportion of the sample at risk of inadequate or excess intake depends on the relative position of the dietary guideline and the median of the distribution. In our samples, the median intake of iron, calcium, energy, and cholesterol is less than the dietary guidelines. Hence, reducing the variation of the distribution of average intake increases the proportion of the sample whose average intake is less than the guidelines (for example, for iron, calcium, and energy) and reduces the proportion of the sample whose average intake is greater than the guidelines (for example, for cholesterol). Conversely, the median intake of protein, vitamin C, and saturated fat lies above the dietary guideline. Hence, reducing the variation in the distribution of average intake reduces the proportion of the sample whose intake is less than the dietary guidelines (for example, for protein and vitamin C) and increases the proportion of the sample whose intake is greater than the dietary guideline (for example, for saturated fat).

Since the purpose of this report is not to provide estimates of the proportion of the *population* at nutritional risk, we do not use weighted data. All distributions presented in this report are estimated with unweighted data. Fraker, Long, and Post (1990) discuss the procedures for estimating the weighted distributions of usual dietary intake.

VI. ESTIMATES OF THE DISTRIBUTIONS OF USUAL DIETARY INTAKE DERIVED FROM SUBSAMPLES OF THE DATA

In the previous chapter we illustrated the importance of adjusting for intraindividual variation in dietary intake. But doing so requires at least two days of intake data from some persons in the sample. This chapter investigates whether "good" estimates of the distribution of usual intake can be derived with fewer days of intake data from at least some persons in the sample. If so, then estimates of the proportion of the population at nutritional risk could be made from smaller-scale food consumption surveys. Reducing the number of days of intake data collected from at least some persons in the sample would significantly reduce the cost of data collection.

We estimated distributions of usual dietary intake using four separate subsamples of each of the two samples of four days of intake data from 600 women and 600 children, respectively. We used the NRC adjustment procedure to remove intraindividual variation from each distribution. We judge the quality of these estimates by comparing them with our "best" estimates of the distribution of usual intake from the full sample presented in the previous chapter. We use three characteristics of the distribution in the comparison: (1) the estimated mean usual intake, (2) the estimated standard deviation of usual intake, and (3) the proportion of the population whose usual dietary intake is below the RDA or above the guideline for maximum intake.

A. CHOOSING THE SUBSAMPLES

Each subsample contains at least one day of data from each of the 600 persons in the sample. We constructed the subsamples by reducing the number of days of intake data available from at least some persons in the sample. The subsamples fall into two categories: (1) subsamples constructed by reducing the number of days of data from ***each*** person in the sample, and (2) subsamples constructed from four days of intake data only from a ***subset*** of the persons in the sample and one day of data from the remainder of the persons in the sample.

Two of the subsamples fall into the first category:

1. Three days of intake data available from each person. (Number of observations = 3 x 600 = 1,800.)

2. Two days of intake data available from each person. (Number of observations = 2 x 600 = 1,200.)

In these subsamples, it is assumed that only two or three days of dietary intake data are available for each person. The days of data included in the sample are the first two or three days reported in the full sample.[1] The days were selected in this manner rather than randomly, to approximate the data available from a survey in which fewer interviews are conducted.

Two subsamples fall into the second category:

1. Four days of data available from 400 persons and one day of data available from the remaining 200 persons. (Number of observations = 400 x 4 + 200 x 1 = 1,800.)

2. Four days of data available from 200 persons and one day of data available from the remaining 400 persons. (Number of observations = 200 x 4 + 400 x 1 = 1,200.)

In these subsamples, the four days of intake data are available only from some persons in the sample, and only the first day of intake data is available for the remainder of the sample. The persons from whom four days of dietary intake data were assumed to be available were selected randomly from the full sample of 600 persons.[2] Each subsample was selected independently of the other subsamples.

‘These may not be the first two or three days of intake data reported by the person in the survey: when the number of days of intake data available from each person exceeded four, the first reported day of intake was included in the MPR analysis file, but the remaining three days of intake data were selected randomly from the remaining days of intake data.

[2]A random number generator available in the computer package SAS created the random samples.

B. THE METHODOLOGY USED TO ESTIMATE THE DISTRIBUTIONS OF USUAL DIETARY INTAKE

To estimate the distributions of usual dietary intake with the subsamples of two and three days of intake data from each person in the sample, we used the same procedures as we used to estimate the distributions from the full sample. For each procedure, we assumed that only the subsample of data was available. For example, we used only the subsample of data to reestimate the interview effects. Similarly, we used the subsample of data to reestimate the adjustment factors.

The procedures used to estimate the distributions of usual dietary intake with the subsamples that contained multiple days of intake data only from some persons in the sample differed in one respect from the procedures used to estimate the other subsamples: the adjustment factors were estimated with data only from the subset of persons for whom we had multiple days of intake data. For example, in the subsample of four days of intake data from 400 persons and one day of intake data from 200 persons (1,800 observations) we used only data from the 400 persons with four days of intake data (1,600 observations) to estimate the adjustment factors. We did so because observations only on one day of intake data do not provide information on the amount of variation of daily dietary intake due either to intraindividual or to interindividual variation. Adding the observations on the intake of persons with only one day of intake data would have increased the complexity of the computations[3] but would not have affected our estimates of the variance ratio. The adjustment factor is applied to the average dietary intake of everyone in the subsample.

C. ESTIMATES OF THE DISTRIBUTIONS DERIVED FROM SAMPLES WITH FEWER DAYS OF INTAKE DATA FROM EACH PERSON

Table VT.1 presents summary statistics of the distributions of usual dietary intake estimated with two and three days of intake data from each person in the sample. For comparison, Table VI.1 also

[3]The ANOVA package in SAS requires that the data set be "balanced"--that is, that the same number of observations be available on each person. This would not be the case if every observation in the subsample were included.

TABLE VI.1

ESTIMATED DISTRIBUTIONS OF USUAL DIETARY INTAKE,
BY NUMBER OF DAYS OF DATA ON EACH INDIVIDUAL

Dietary Component	Statistic	Number of Days of Data Reported by Each Person		
		Four	Three	Two
Number of Observations		**2,400**	1,800	1,200
Sample of 600 Women				
Vitamin C	Mean	73.6	73.1	74.6
	Standard deviation	37.9	34.9	32.6
	Percentage below RDA	45.3	44.2	41.3
Calcium	Mean	548.6	547.8	551.1
	Standard deviation	236.0	241.0	249.1
	Percentage below RDA	86.5	87.0	87.5
Iron	Mean	9.7	9.7	9.8
	Standard deviation	3.0	2.9	3.1
	Percentage below RDA	94.8	95.3	94.2
Protein	Mean	57.7	57.7	58.6
	Standard deviation	15.8	16.0	16.7
	Percentage below RDA	0.3	0.3	0.2
Energy	Mean	**1,473.7**	1484.8	1,513.0
	Standard deviation	418.0	410.0	377.3
	Percentage below RDA	94.3	94.5	95.5
Cholesterol	Mean	279.5	280.1	286.9
	Standard deviation	112.0	104.8	119.2
	Percentage above guideline	34.8	33.7	36.3
Saturated Fat	Mean	12.6	12.7	12.7
	Standard deviation	1.9	1.9	1.9
	Percentage above guideline	92.5	91.7	94.3
Sample **of 600 Children**				
Vitamin C	Mean	80.6	80.8	**83.0**
	Standard deviation	34.7	31.9	**36.8**
	Percentage below RDA	11.0	7.3	**8.0**
Calcium	Mean	761.2	763.7	**760.5**
	Standard deviation	222.9	230.7	**223.9**
	Percentage below RDA	62.8	61.7	**61.8**

TABLE VI.1 (continued)

Dietary Component	Statistic	Number of Days of Data Reported by Each Person		
		Four	Three	Two
Iron	Mean	9.8	9.9	10.0
	Standard deviation	3.1	3.3	3.4
	Percentage below RDA	65.3	64.3	62.7
Protein	Mean	52.9	53.0	53.5
	Standard deviation	12.0	11.8	12.7
	Percentage below RDA	0.0	0.0	0.0
Energy	Mean	1,403.4	1,402.8	1,413.6
	Standard deviation	304.5	293.5	295.0
	Percentage below RDA	91.2	91.8	91.2
Cholesterol	Mean	250.3	249.8	252.2
	Standard deviation	80.4	78.6	97.3
	Percentage above guideline	20.8	21.3	25.3
Saturated Fat	Mean	13.8	13.9	13.9
	Standard deviation	1.5	1.6	1.9
	Percentage above guideline	99.1	99.8	99.8

SOURCE: Random samples taken from the MPR analysis file that contains four days of intake data from the core and low-income samples of the 1985 CSFII.

presents summary statistics of the distributions of usual dietary intake estimated with the full sample of four days of intake data from all 600 persons in the sample. The full sample contains a total of 2,400 observations on daily dietary intake, the subsample of three days of intake data from each person contains 1,800 observations, and the subsample of two days of intake data from each person contains 1,200 observations.

The means of the distributions of usual dietary intake estimated from three days of intake data are similar to those estimated from four days of intake data, and are not systematically higher or lower than those estimated from the full sample. However, the mean of 13 of the 1.4 distributions estimated with two days of intake data is higher than the mean estimated with the full sample of data. This difference may be due to persons' reporting a higher intake of dietary components in the first interview: the proportion of the observations on daily dietary intake from the first interview is higher in the sample with two days of intake data from each person than in the full sample. However, the difference in the estimated mean is small; in each distribution, the difference is much smaller than the standard deviation of the distribution. The difference in the estimated means is largest in the distribution of vitamin C intake by children, but the difference is still less than 3 percent.

Whether the estimate of the mean made with the smaller sample is better or worse than the estimate made with the full sample depends on the reason for the higher consumption reported in the first interview. If in the first interview the respondent gives an accurate record of food consumption but underreports his or her consumption in later interviews, the estimate of the mean with two days of intake data may be better than the estimate derived from four days of intake data. If, on the other hand, the respondent exaggerates food consumption in the first interview, the estimate of the mean made from the full sample will be a better estimate than the estimate made from the smaller sample.

The standard deviations of the distributions of usual dietary intake estimated with three days of intake data do not vary systematically from the standard deviation of the distributions estimated with the full sample. However, the standard deviation of the distributions of usual dietary intake of 10 of the 14 dietary components estimated with the sample of only two days of intake data are larger than the standard deviations of the distributions estimated from the full sample. A possible explanation is that with only two days of intake data less of the intraindividual variation can be identified and removed.

The estimates of the proportion of the sample whose usual dietary intake is either below the RDA or above the maximum dietary guideline that are derived from a sample with only three days of intake data are similar to those derived from the full sample. The three-day estimates of the proportion of the sample of women at nutritional risk from a deficient or excess intake of all seven dietary components differed by less than 2.5 percentage points from the estimates made with the full sample. Similarly, the estimates of the proportion of the sample of children at nutritional risk from a deficient or excess intake of all dietary components except vitamin C differed by less than 2.5 percentage points from the estimates made with the full sample. Of the 14 distributions estimated, the difference between the two estimates varied by more than 1 percentage point in only four cases, and the difference was only 1.1 percentage points in three of the four cases.

The two-day estimates of the proportion of the sample at nutritional risk differ from the estimates made with the full sample by more than the three-day estimates. The estimates of the proportion of the sample at nutritional risk differ by more than 2.5 percentage points from the estimates made with the full sample in four distributions--vitamin C intake by women and vitamin C, iron, and cholesterol intake by children. The two-day estimates differ from the four-day estimates by more than 1 percentage point in 7 of the 14 distributions.

The difference across subsamples in the estimates of the proportion of the sample at nutritional risk is due to differences in the estimates of both the mean and the standard deviation. Hence, even if the estimate of the standard deviation of usual dietary intake is constant across subsamples, differences in the proportion of the sample at nutritional risk would occur because of differences in estimated means. The difference in means across subsamples is purely a result of using different subsamples, and is not a result of applying different NRC adjustment factors.

D. ESTIMATES OF DISTRIBUTIONS DERIVED FROM SAMPLES WITH MULTIPLE DAYS OF DATA FROM ONLY A SUBSET OF PERSONS

Table VI.2 presents summary statistics of two distributions of usual dietary intake estimated with four days of intake data from 200 and 400 persons, respectively, and one day of data from the remainder of the sample. For comparison, the table also presents summary statistics of the distribution of usual dietary intake estimated with the full sample of four days of data from everyone in the sample. The full sample contains 2,400 observations on daily dietary intake:, the sample of four days of data from 400 persons and one day of data from 200 persons contains 1,800 observations, and the sample of four days of data from 200 persons and one day of data from 400 persons contains 1,200 observations.

The estimated means of the distributions with four days of data from 400 persons do not differ systematically from the estimated means with the full sample of four days of data from 600 persons. The difference between the estimates is small--less than 2.5 percent in all distributions.

However, the means of all distributions estimated with four days of data from only 200 persons in the sample are at least as large as the means estimated with the full sample. As we found in Table VI.1, the estimated means are higher when the sample contains a higher proportion of observations on daily dietary intake that were reported on the first day of interviews. The largest difference pertains to the estimated mean intake of iron by women, which is 5 percent higher when estimated with four days of data from only 200 persons, rather than with the full sample. In the sample of children, the difference between the means of the distributions estimated with the

TABLE VI.2

ESTIMATED DISTRIBUTIONS OF USUAL DIETARY INTAKE, BY THE NUMBER OF PERSONS REPORTING FOUR DAYS OF DATA

Dietary Component	Statistic	Number of Individuals Reporting Four Days of Data (Remaining Individuals Report One Day) 600	400	200
Number of Observations		**2,400**	1,800	1,200
Sample of 600 Women				
Vitamin C	Mean	**73.6**	71.9	76.2
	Standard deviation	**37.9**	34.8	38.7
	Percentage below RDA	**45.3**	45.3	40.8
Calcium	Mean	**548.6**	550.7	564.0
	Standard deviation	**236.0**	241.3	287.4
	Percentage below RDA	**86.5**	86.8	84.3
Iron	Mean	**9.7**	9.8	10.2
	Standard deviation	**3.0**	3.3	3.4
	Percentage below RDA	**94.8**	94.9	93.0
Protein	Mean	**57.7**	57.8	60.2
	Standard deviation	**15.8**	16.5	20.8
	Percentage below RDA	**0.3**	0.3	0.3
Energy	Mean	**1,473.7**	**1,474.4**	**1,538.6**
	Standard deviation	**418.0**	415.7	497.0
	Percentage below RDA	**94.3**	94.7	89.8
Cholesterol	Mean	**279.5**	277.5	294.0
	Standard deviation	**112.0**	114.3	148.1
	Percentage above guideline	**34.8**	33.6	36.3
Saturated Fat	Mean	**12.6**	12.5	12.6
	Standard deviation	**1.9**	1.9	2.4
	Percentage above guideline	**92.5**	91.6	87.0
Sample of 600 Children				
Vitamin C	Mean	**80.6**	80.5	81.5
	Standard deviation	**32.0**	35.7	39.6
	Percentage below RDA	**11.0**	14.2	15.3
Calcium	Mean	**761.2**	760.1	766.5
	Standard deviation	**222.9**	233.4	237.1
	Percentage below RDA	**62.8**	62.7	58.7
Iron	Mean	**9.8**	9.9	9.9
	Standard deviation	**3.1**	3.3	2.5
	Percentage below RDA	**65.3**	65.7	**60.6**

TABLE VI.2 (continued)

Dietary Component	Statistic	Number of Individuals Reporting Four Days of Data (Remaining Individuals Report One Day)		
		600	400	200
Protein	Mean	52.9	53.2	53.5
	Standard deviation	12.0	12.6	13.1
	Percentage below RDA	0.0	0.5	0.6
Energy	Mean	1,403.4	1,403.0	1409.0
	Standard deviation	304.5	314.7	310.3
	Percentage below RDA	91.2	90.7	91.4
Cholesterol	Mean	250.3	249.2	254.7
	Standard deviation	80.4	97.4	106.3
	Percentage above guideline	20.8	23.7	27.8
Saturated Fat	Mean	13.8	13.8	13.8
	Standard deviation	1.5	1.7	1.8
	Percentage above euideline	99.7	98.8	98.0

SOURCE: Random samples taken from the MPR analysis file that contains four days of intake data from the core and low-income samples of the 1985 CSFII.

subsample of four days of intake data on 200 persons and with the full sample is never more than 2 percent.

In nearly all distributions, the standard deviations of the distributions of usual dietary intake estimated with four days of intake data from 400 persons and four days of intake data from 200 persons are lower than the standard deviations of the distributions estimated with the full sample. In 11 of the 14 distributions estimated with the sample of four days of data from 400 persons, the standard deviations exceeded the standard deviations of the distributions estimated with the full sample. In all but one of the distributions estimated with four days of data from 200 persons, the standard deviations exceeded the standard deviations estimated with the full sample.

Estimates of the proportion of the population at nutritional risk that are derived from the sample that contains four days of intake data from 400 persons were similar to those derived from the full sample. All estimates of the proportion of the sample of women at nutritional risk were within 2.5 percentage points of the estimates derived from the full sample, and, with the exception of the estimate of the proportion of the sample at risk from an excess intake of cholesterol, all estimates were within 1 percentage point of the estimates derived from the full sample. All estimates of the proportion of the sample of children at nutritional risk were within 1 percentage point of the estimates derived from the full sample, with the exception of the estimates of the proportion of the sample at risk from a deficient or an excess intake of vitamin C and cholesterol.

Estimates of the proportion of the sample at nutritional risk derived from four days of intake data on only 200 persons were less similar to the estimates derived from the full sample. For the sample of women, only the estimates of the proportion of the sample at risk from a deficient intake of calcium, iron, and protein and an excess intake of cholesterol were within 2.5 percentage points of the estimates made with the full sample; and only the estimate of the proportion of the sample at risk from a deficient intake of protein was within 1 percentage point of the estimate derived from the full sample. Similarly, for the sample of children, only the estimates of the proportion of the sample at nutritional risk from a deficient or excess intake of protein, energy, and saturated fat that were derived from the sample of four days of intake data on 200 persons were within 2.5 percentage points of the estimates derived from the full sample. Only the estimates of the proportion of the population at risk from a deficient intake of protein and energy were within 1 percentage point of the estimates derived from the full sample.

VII. THE IMPLICATIONS OF THE ESTIMATES FOR THE DESIGN OF FOOD CONSUMPTION SURVEYS

This chapter discusses the implications of the estimates presented in the previous two chapters for the design of food consumption surveys. Section A discusses the number of days of dietary intake data from each person that is required to derive good estimates of the distributions of usual dietary intake. Section B discusses the number of persons for whom multiple days of dietary intake data must be collected to derive good estimates of the distributions of usual dietary intake. Section C discusses whether better estimates can be obtained from a small number of days of intake data from everyone in the sample or from a larger number of days of intake data only from some of the persons in the sample and a single day of data from the remaining persons in the sample. Section D discusses the limitations of the study.

To summarize the results of our comparisons of estimates of the distributions of usual dietary intake, Tables VII.1 and VII.2 indicate for each subsample the estimates of the proportion of the sample at nutritional risk that differ by more than a specified amount from the best estimates derived from the full sample. Table VII.1 indicates those estimates that differ by more than 2.5 percentage points from the best estimates. Table VII.2 indicates those estimates that differ by more than 1 percentage point from the best estimates. It also indicates the estimates of the proportion of the sample at nutritional risk made without adjusting for intraindividual variation that vary by more than 2.5 percentage points from our best estimates. Table VII.2 is identical to Table WI.l, with the exception that it indicates estimates that differ from our best estimates by more than 1 percentage point.

The level of precision required for the estimates will depend on their intended use. Our choice of differences of 2.5 and 1 percentage points as criteria for judging the estimates of the proportion of the sample at nutritional risk was somewhat arbitrary. It should be noted that although 2.5 percentage points is an acceptable level of error for many statistical applications, an error of this

TABLE VII.1

ESTIMATES OF THE PROPORTION OF THE SAMPLE AT NUTRITIONAL RISK THAT DIFFER FROM OUR BEST ESTIMATES BY MORE THAN 2.5 PERCENT

	Unadjusted Distribution Full Sample: 4 days from 600 Persons	3 Days from 600 Persons	2 Days from 600 Persons	4 Days from 400 Persons 1 Day from 200 Persons	4 Days from 200 Persons 1 Day from 400 Persons
Number of Observations	2,400	1,800	1,200	1,800	1,200
Sample of Women					
Vitamin C	x		x		x
Calcium					
Iron	x				
Protein					
Energy	x				x
Cholesterol					
Saturated Fat	x				x
Sample of Children					
Vitamin C	x	x	x		x
Calcium					x
Iron			x		x
Protein					
Energy	x				
Cholesterol	x		x	x	x
Saturated Fat	x				

SOURCE: Tables VI.1 and VI.2.

TABLE VII.2

ESTIMATES OF THE PROPORTION OF THE SAMPLE AT NUTRITIONAL RISK THAT DIFFER FROM OUR BEST ESTIMATES BY MORE THAN 1 PERCENT

	Unadjusted Distribution Full Sample: 4 Days from 600 Persons	3 Days from 600 Persons	2 Days from 600 Persons	4 Days from 400 Persons 1 Day from 200 Persons	4 Days from 200 Persons, 1 Day from 400 Persons
Number of Observations	**2,400**	**1,800**	1,200	1,800	**1,200**
Sample of Women					
Vitamin C	x	x	x		x
Calcium	x				x
Iron	x				x
Protein					
Energy	x		x		x
Cholesterol	x	x	x	x	x
Saturated Fat	x		x		x
Sample of Children					
Vitamin C	x	x			x
Calcium	x	x			x
Iron					x
Protein					
Energy	x				
Cholesterol	x		x	x	x
Saturated Fat	x				x

SOURCE: **Tables VI.1 and VI.2.**

magnitude could have important implications in designing policy affecting large populations. Even a 1 percentage point difference in an estimate of the proportion of people at nutritional risk could account for about 2.5 million persons if applied to the U.S. population as a whole.'

The "worst" estimates--those that differ most widely from our best estimates--are those derived from the full sample of data but without the NRC adjustment procedure. The estimate of the proportion of the sample at nutritional risk for 8 of the 14 distributions differs from our best estimate by more than 2.5 percentage points. Even when we estimated the proportion. of the sample at nutritional risk with alternative samples of only 1,200 observations, the estimates were typically closer to the best estimates after the adjustment for intraindividual than were those derived from the full 2,400 observations without the NRC adjustment procedure.

The robustness of the estimates of the proportion of the sample at nutritional risk varies by dietary component. Estimates of the proportion of women at risk of a deficient intake of calcium, iron, protein, and cholesterol are robust to changes in the sample, as are estimates of the proportion of children at nutritional risk of a deficient or excess intake of protein, energy, and saturated fat. Conversely, estimates of the proportion of the samples of both women and children at nutritional risk of a deficient intake of vitamin C, and the proportion of the sample of children at nutritional risk of an excess intake of cholesterol, vary across samples. It is interesting that the most robust estimates are made from roughly symmetric distributions, while the distributions of vitamin C intake by women and by children and cholesterol intake by children are three of the least symmetric distributions.

---

[1]**A** reviewer pointed out that the appropriate criteria used to measure the accuracy of the estimates may vary with the size of the estimate. Thus, the criteria used to judge the precision of the estimate of the proportion of the population with food energy intake below the RDA (which is over 90 percent) would be different than the criteria used to estimate the proportion of the population with protein intake below the RDA (which is less than 1 percent).

A. HOW MANY DAYS OF DIETARY INTAKE DATA ARE REQUIRED FROM EACH PERSON IN THE SAMPLE?

Our results suggest that fairly good estimates can be made with only three days of intake data from everyone in the sample. Most of the estimates of the proportion of the sample at nutritional risk are within 1 percentage point of our best estimates. Similarly, the means and standard deviations of the estimated distributions closely approximate the means and standard deviations of the distributions estimated with the full sample.

Some estimates made with only two days of intake data from everyone in the sample differ more substantially from our best estimates. Although only estimates of the proportion of the sample at nutritional risk made from 4 of the 14 distributions differ by more than 2.5 percentage points from our best estimates, estimates for half of the distributions vary by more than 1 percentage point from our best estimate.

These result suggest that the returns from collecting an extra day of data diminish as more days of data are collected from each person in the sample. While our estimates derived from three days of data improve upon those derived from only two days of intake data, the estimates derived from four days of intake data are only slightly better than those derived from three days of intake data. This result confirms the previous findings of Ritenbaugh et al. (1988).

As indicated in Table VI.1, the mean of the usual intake estimated with two days of data from each person in the sample is nearly always higher than the mean estimated with four days of data from each person in the sample, This result may be due to the fact that persons report a higher intake of food in the first interview. Hence, the higher proportion of dietary intake reported in the first interview in the smaller sample increases the estimate of the mean.

A higher estimate of the mean of the distribution of usual dietary intake decreases the estimate of the proportion of the sample at risk of a deficient intake of a dietary component and increases the estimate of the proportion of the sample at risk of an excessive intake of a dietary component. Hence, to some degree, the difference in the estimates of the proportion of the sample at nutritional risk across samples is due to differences in estimated means rather than differences in the NRC adjustment factor.

B. FROM HOW MANY PERSONS IN THE SAMPLE ARE MULTIPLE DAYS OF INTAKE DATA REQUIRED?

Our results suggest that good estimates can be made by using four days of intake data from 400 persons in the sample and one day of data from the remainder of the sample. Only estimates of the proportion of the sample at nutritional risk that are derived from the distributions of cholesterol intake by women and vitamin C and cholesterol intake by children differ by more than 1 percentage point from our best estimate.

However, good estimates of the proportion of the sample at nutritional risk cannot always be derived from four days of intake data from only 200 persons in the sample and one day of data from the remainder of the sample. Half of the estimates of the proportion of the sample at nutritional risk differ by more than 2.5 percentage points from our best estimates. Only three estimates of the proportion of the sample at nutritional risk are less than 1 percentage point from our best estimates.

C. ARE BETTER ESTIMATES OBTAINED FROM A SMALL NUMBER OF DAYS OF INTAKE DATA FROM EVERYONE IN THE SAMPLE OR FROM A LARGER NUMBER OF DAYS OF INTAKE DATA FROM A SUBSET OF THE PERSONS IN THE SAMPLE?

We can address this question by comparing the estimates derived from three days of intake data from all 600 persons in the sample with those derived from four days of intake data from 400 persons and one day of intake data from 200 persons because both samples contain a total of 1,800 observations. Similarly, we can compare the estimates derived from two days of intake data from all 600 persons in the sample with those derived from four days of intake data from 200 persons and one day of intake data from 400 persons because both samples contain a total of 1,200 observations.

Estimates derived from three days of intake data from 600 persons are similar to those derived from four days of intake data from 400 persons and one day of intake data from 200 persons--both the means and the standard deviations of the estimated distributions are similar, as are the estimates of the proportion of the sample at nutritional risk made from each subsample. Only one estimate of the proportion of the sample at nutritional risk derived from the sample of three days of intake data from 600 persons differs from our best estimate by more than 2.5 percentage points, compared with two estimates derived from the sample of four days of data from 400 persons and one day of data from 200 persons.

Our results suggest that estimates derived from two days of data from everyone in the sample are closer to the best estimates than are those derived from four days of data from only 200 persons and one day from the remainder of the sample. Estimates of the proportion of the sample at nutritional risk that are derived from two days of data from all 600 persons in the sample differ from our best estimates by more than 2.5 percentage points in only four distributions; estimates of the proportion of the sample at nutritional risk derived from four days of data from 200 persons in the sample and one day of data from 400 persons differ from our best estimates by more than 2.5 percentage points in seven distributions. Similarly, more estimates derived from two days of data on 600 persons are within 1 percentage point of our best estimate than estimates from four days of data on 200 persons and one day of data on 400 persons.

These results suggest that food consumption surveys should be designed to collect multiple days of intake data from each person rather than from just a subset of the persons in the sample. If the cost of each interview is equal, equally good or better estimates of the proportion of the sample at nutritional risk can be obtained at the same cost by collecting multiple days of intake data from everyone in the sample rather than from just a subset of the persons in the sample.

Past food consumption surveys have found that the response rate falls in later waves of interviews. Hence, a fourth interview is, in effect, more expensive than the second and third interview because it requires a larger initial sample to collect a sample with a complete set of intake data on each person. Thus collecting two or three days of intake data from everyone in the sample may also be less expensive than collecting four days of intake data from a subset of the persons in the sample.

D. LIMITATIONS OF OUR STUDY

Two limitations of this study should be kept in mind in an interpretation of its tindings. First, the study estimates distributions based on subsamples of data only for seven dietary components in two separate samples. Hence, the results may not be generalizable to estimates of th.e distributions of the usual intake of other dietary components by other groups of individuals.

Second, we compare our estimates derived from subsamples of the data with *estimates* of the distribution derived by applying the NRC adjustment factor to the full sample, rather than with the *true* distribution. Hence, if we had more than four days of intake data from each person in the sample, our "best" estimates may change. Furthermore, if the assumptions underlying the *NRC* adjustment procedure are not met, the full sample estimates may not be the *best* estimates of the distributions of usual dietary intake. For example, if respondents accurately report food consumption in the first interview and underreport food consumption in subsequent interviews, an estimate of the mean of the true distribution using data only from the first interview may be more accurate than an estimate made with the full sample. To the extent that our full sample estimates differ from the true distributions, the usefulness of the comparisons made in this study to the design of food consumption surveys is limited.

The second possible limitation of our study could be addressed in two ways. First, we could repeat the analysis discussed in this report using "fake" data. These fake data could be constructed by randomly generating "dietary intake data" from a normal distribution with a known mean and standard deviation. The advantages of using constructed data are that the true distributions of usual dietary intake would be known and that the data would be generated so that they automatically satisfy the assumptions underlying the NRC adjustment procedure. We could then compare estimates of the distributions of usual dietary intake made using subsamples of the constructed data with the true distributions of usual dietary intake.

Alternatively, we could repeat the analysis using data that include many days of data from each person in the sample. As an example, food consumption data for 365 days were collected from a group of 29 volunteers in Beltsville, Maryland (Basiotis et al., 1987). We could then compare estimates of the distributions of usual dietary intake made using subsamples of those data with the estimates made using all 365 days of data from each person. The advantage of this strategy is that the estimates made using 365 days of data from each person would be closer to the "true" distribution than those in this study made using only four days of data.

REFERENCES

Aicken, Michel. "An Application of Unmixing to Vitamin A Intake Data." Final report under USDA cooperative agreement #58-3198-6-62, Tucson, AZ: The University of Arizona, July 1988.

Basiotis, P. Peter, Susan 0. Welsh, Frances J. Cronin, June L. Kelsay, and Walter Mertz. "Number of Days of Food Intake Records Required to Estimate Individual and Group Nutrient Intakes with Defined Confidence." ***The Journal of Nutrition, vol. 117,*** September 1987, pp. 1638-1641.

Battese, G.E., S.M. Nusser, and W.A. Fuller. 'Estimation of the Distribution of Usual Intakes for Selected Dietary Components Using Data From the 1985-86 Continuing Survey of Food Intakes by Individuals." Ames, IA: Center for Agricultural and Rural Development and Department of Statistics, Iowa State University, 1988.

Fraker, Thomas M., Sharon K. Long, and Charles E. Post. "Analyses of the 1985 Continuing Survey of Food Intakes by Individuals." Washington, D.C.: Mathematica Policy Research, September 1990.

Fraker, Thomas M., and Charles Post. "Documentation for a SAS Analysis File Containing Merged 4-Day Data for the Core and Low-Income Samples of the 1985 CSFII." Washington, D.C.: Mathematica Policy Research, March 1989.

Johnson, Stanley R. et al. "Methodological Factors Affecting the Measurement of Dietary Status." Ames, IA: Center for Agricultural and Rural Development and Department of Statistics, Iowa State University, 1988.

Liu, Kiang, Jeremiah Stamler, Alan Dyer, Jeffrey McKeever, and Patricia McKeever. "Statistical Methods to Assess and Minimize the Role of Intra-Individual Variability in Obscuring the Relationship between Dietary Lipids and Serum Cholesterol." ***Journal of Chronic Diseases,*** vol. 31, 1978, pp. 399-418.

Maddala, G.S. Econometrics. Tokyo, Japan: McGraw-Hill Kogakusha, 1979.

Maddala, G.S. "The Use of Variance Components in Pooling Cross Section and Time Series Data." *Econometrica*, vol. 39, 1971, pp. 341-358.

National Research Council, Subcommittee on Criteria for Dietary Evaluation. ***Nutrient Adequacy: Assessment Using Food Consumption Surveys.*** Washington, D.C.: National Academy Press, 1986.

National Research Council, Committee on Dietary Allowances. ***Recommended Dietary Allowances.*** Washington D.C.: National Academy Press, 1989.

Pao, E.M., KS. Sykes, and Y.S. Cypel. ***USDA Methodological Research for Large-Scale Dietary Intake Surveys, 1975-88.*** Hyattsville, MD: Human Nutrition Information Service, U.S. Department of Agriculture, 1989.

Ritenbaugh, Cheryl, et al. "Methodological Issues in Food Consumption Surveys." Final report under USDA cooperative Agreement #58-3198-6-62, Tucson, AZ: The University of Arizona, July 1988.

Sempos, Christopher T., Nancy E. Johnson, Everett L. Smith, and Catherine Gilligan. "Effects of Intraindividual and Interindividual Variation in Repeated Dietary Records." American *Journal of Epidemiology*, vol. 121, no. 1, 1985.

APPENDIX A

DETAILED TECHNICAL DESCRIPTION OF THE NRC ADJUSTMENT PROCEDURE

This appendix provides a detailed technical description of the derivation of the NRC adjustment procedure for estimating distributions of usual dietary intake.[1] Section A presents the notation used in this appendix. Section B lists the assumptions underlying the procedure. Section C describes the derivation of the formula for the estimated distribution of usual dietary intake. Section D describes the estimation of intraindividual and interindividual variation.

A. NOTATION

We denote the daily dietary intake of individual i on day t by y_{it}, and the usual dietary intake of individual i by x_i. T denotes the number of days of data available for each individual, and N denotes the number of individuals in the sample. The average dietary intake of individual i is denoted by y_i, where:

$$y_i = \frac{\sum_t y_{it}}{T}. \tag{1}$$

We denote the overall average dietary intake, or grand mean, by y, where:

$$y = \frac{\sum_i y_i}{N} = \frac{\sum_i \sum_t y_{it}}{NT}. \tag{2}$$

Similarly, we denote the overall average usual dietary intake by **x**, where:

$$x = \frac{\sum_i x_i}{N}. \tag{3}$$

B. ASSUMPTIONS UNDERLYING THE NRC ADJUSTMENT PROCEDURE

The NRC adjustment procedure is based on the following two assumptions:

---

[1] The contents of this appendix are drawn from private correspondence with Professor George Beaton.

Assumption 1. The daily dietary intake of individual i on day t, y_{it}, can be decomposed into the sum of the usual dietary intake of person i, x_i, and an error, e_{it}:

(4)
$$y_{it} = x_i + e_{it}.$$

Four assumptions are made about the error:

1.1 The mean value of the error is zero.

1.2 The error is uncorrelated with the usual dietary intake, x_i.

1.3 The error does not exhibit serial correlation. That is, it is uncorrelated with the error on previous or subsequent days.

1.4 The error is homoskedastic. That is, it has a constant variance.

Assumption 2. The usual dietary intake, x_i, is normally distributed with a mean of x and a standard deviation of $sd(x_i)$. Thus, we can write x_i as:

(5)
$$x_i = x + \beta_i \cdot sd(x_i),$$

where β_i measures the distance of the average dietary intake of person i from the mean value. The average dietary intake, y_i, is also distributed as a normal with a mean of y and a standard deviation of $sd(y_i)$. Thus, we can write y_i as:

(6)
$$y_i = y + \alpha_i \cdot sd(y_i),$$

and it is assumed that $\alpha_i = \beta_i$

C. DERIVATION OF THE FORMULA FOR THE DISTRIBUTION OF USUAL DIETARY INTAKE

We cannot observe the distribution of usual dietary intake given in equation (5). But, as shown in this section, we can observe β_i and the mean dietary intake, x, in terms of observable dietary intake. From assumption 1, we know that the mean value of the distribution of usual dietary intake is equal to the mean value of average dietary intake. This can be shown by noting from equation (4) that:

(7)
$$x_i = y_{it} - e_{it},$$

and summing both sides of the equation over the number of days of data available for each individual and dividing by T yields:

(8)
$$x_i = \frac{\sum_t y_{it}}{T} - \frac{\sum_t e_{it}}{T} = y_i - \frac{\sum_t e_{it}}{T}$$

and the definition of x in equation (3) and using assumption 1.1 to set the average value of the error equal to zero yield:

(9)
$$x = \frac{\sum_i x_i}{N} = \frac{\sum_i y_i}{N} - \frac{\sum_i \sum_t e_{it}}{NT} = \frac{\sum_i y_i}{N} = y.$$

By assumption 2, β_i is equal to α_i, and β_i can thus be derived in terms of average observable dietary intake by rearranging equation (6) to yield:

(10)
$$\beta_i = \alpha_i = \frac{y_i - y}{sd(y_i)}.$$

Taking equation (5) and substituting y for x and replacing β_i by equation (10) yield,

(11)
$$x_i = y + (y_i - y) \cdot \frac{sd(x_i)}{sd(y_i)}.$$

Note that the distribution of usual dietary intake, x_i, is centered around the overall mean of the average dietary intake, y. But, because the term $\frac{sd(x_i)}{sd(y_i)}$ is less than 1, each point in the distribution of usual dietary intake is nearer the mean than is the corresponding point in the distribution of average dietary intake. The term $\frac{sd(x_i)}{sd(y_i)}$ is called the "adjustment factor," or "attenuation factor."

It is sometimes written as the square root of the ratio of the variance of usual dietary intake, $V(x_i)$, to the variance of average dietary intake, $V(y_i)$:

(12)
$$\textit{Adjustment Factor} = \frac{sd(x_i)}{sd(y_i)} = \sqrt{\frac{V(x_i)}{V(y_i)}}.$$

It is helpful to write the adjustment factor in terms of inter-individual variation and intraindividual variation. Interindividual variation is measured by the variance of usual dietary intake, $V(x_i)$, and intraindividual variation is measured by the variance of the error, $V(e_{it})$. Taking equation (4) and summing over the number of days of data available for each individual and dividing by T yields:

(13)
$$y_i = x_i + \frac{\sum_t e_{it}}{T}$$

and taking the variance of both sides of the equation yields:

(14)
$$V(y_i) = V(x_i) + \frac{V(e_{it})}{T}.$$

Substituting equation (14) into the formula for the adjustment factor, equation (12), yields:

(15)
$$\textit{Adjustment Factor} = \sqrt{\frac{V(x_i)}{V(x_i) + \frac{V(e_{it})}{T}}}.$$

D. ESTIMATING INTRAINDMDUAL AND INTERINDIVIDUAL VARIATION

Both interindividual variation and intraindividual variation are unobservable. But we can estimate interindividual and intraindividual variation from an analysis of variance (ANOVA) regression in which daily dietary intake y_{it} is regressed on N-l dummy variables--the nth dummy variable equals 1 if the intake is by the nth individual. The best estimate of the intraindividual variation, $\hat{V}(e_{it})$, is given by the mean square error (MSERROR) of the ANOVA:

$$\hat{V}(e_{it}) = MSERROR. \tag{16}$$

The best estimate of interindividual variation, $\hat{V}(x_i)$, is given by:

$$\hat{V}(x_i) = \frac{MSMODEL - MSERROR}{T}, \tag{18}$$

where MSMODEL denotes the mean square of the model. It is beyond the scope of this report to show the optimality of these estimators.[2] But a comparison of the formulas for the estimators and the formulas for intraindividual and interindividual variation provides some insight. The formula for the MSERROR:

$$MSERROR = \frac{\sum_i \sum_t (y_{it} - y_i)^2}{NT-1}, \tag{19}$$

is similar to the formula for the intraindividual variation:

$$V(e_{it}) = V(y_{it} - x_i) = \frac{\sum_i \sum_t (y_{it} - x_i)^2}{NT}. \tag{20}$$

---

[2] See Maddala (1971) for a proof.

The MSERROR differs from $V(e_{it})$ only by replacing the unknown usual dietary intake, x_i, by the average dietary intake, y_i, and reducing the degrees of freedom in the denominator by 1. To understand the formula for the interindividual variation, note that:

$$\text{(21)} \qquad MSMODEL = \frac{T}{N-1} \cdot \sum_i (y_i - y)^2,$$

and

$$\text{(22)} \qquad V(x_i) = V(y_i) - \frac{V(e_{it})}{T} = \frac{\sum_i (y_i - y)^2}{N} - \frac{V(e_{it})}{T}.$$

Substituting MSERROR for $V(e_{it})$ and $\frac{MSMODEL}{T}$ for $\frac{\sum_i (y_i - y)^2}{N}$ in equation (22) yields the estimate of interindividual variation given in equation (18). It is possible that the MSERROR of the regression can exceed the MSMODEL, implying a negative value of the variance, $\hat{V}(x_i)$. If the MSERROR does exceed the MSMODEL, the estimated variance is set to zero.

The estimates of intraindividual and interindividual variation can be substituted into the formula for the adjustment factor (equation (15)) to yield an estimate of the distribution of usual dietary intake of:

$$\hat{x}_i = y + (y_i - y) \cdot \text{Estimated Adjwtment Factor},$$

where:

$$\text{Estimated Adjustment Factor} = \sqrt{\frac{MSMODEL - MSERROR}{MSMODEL}}.$$

The computer code for calculating these formulas is given on page A-9 of Fraker, Long, and Post (1990).

APPENDIX B

ALTERNATIVE PROCEDURES FOR ESTIMATING THE DISTRIBUTION OF USUAL DIETARY INTAKE

This appendix discusses two alternative procedures for estimating the distribution of usual dietary intake with only a few days of intake data from each person in a sample. The first alternative approach, proposed by Battese et al. (1988), is a parametric approach: it is based on an assumption about the distribution of usual dietary intake. The second alternative approach, proposed by Aicken (1988), is a completely nonparametric approach: it is not based on any assumption about the distribution of usual dietary intake. The advantage of both procedures is that, unlike the NRC adjustment procedure, the distribution of usual dietary intake need not be distributed as a normal. However, both these approaches have two disadvantages not shared by the NRC adjustment procedure: (1) they are computationally complex, and (2) they require more than two days of intake data from each person.

A. A PARAMETRIC PROCEDURE PROPOSED BY BATTESE ET AL.

The procedure developed by Battese et al. for estimating the distribution of usual intake can be applied to asymmetric distributions. The assumption of the normality of the distribution of usual dietary intake, required for the NRC adjustment procedure, is not required for this procedure. However, the Battese procedure does require that an assumption be made about the distribution of usual dietary intake. As an example of their general procedure, Battese et al. estimated distributions of usual nutrient intake under the assumption that the distributions were either Gamma or Weibull.

In common with the NRC adjustment procedure, the Battese procedure is based on the assumptions that (1) the mean value of the error e_{it} (in equation 11.1) is zero, (2) the error is uncorrelated with the usual intake, and (3) the error is uncorrelated with the error on subsequent and previous days. However, the assumption of a constant variance of the error, required for the NRC adjustment procedure, is replaced by the assumption that the standard deviation of the error and the cube root of the third moment of the error are linear functions of the usual intake, implying that intraindividual variation in dietary intake is greater for those individuals whose usual intake is higher.

The Battese procedure is not based on an adjustment of the distribution of average observed dietary intake. Instead, the parameters of the distribution of usual dietary intake are estimated from the moments (for example, the mean, the standard deviation, and the degree of skewness) of the average dietary intake.

Battese et al. estimated distributions of usual intake for five nutrients: iron, protein, calcium, energy, and vitamin C. These nutrients were chosen because their distributions are positively skewed, and it would thus be inappropriate to estimate the distributions under the assumption of normality. The distributions of the average intake of iron, calcium, and vitamin C were found to resemble the Gamma distribution; and the distributions of the average intake of protein and energy were found to resemble the Weibull distribution.

The estimated distributions of usual nutrient intake were found to be more symmetric than the distributions of average nutrient intake. Thus, the degree of skewness, in addition to the variance, differs between the distributions of average and usual intake. This difference highlights the difference between the Battese procedure and the NRC adjustment procedure--the NRC procedure preserves the skewness of the distribution of average intake in its estimate of the distribution of usual intake while the Battese procedure allows the shape of the estimated distribution of usual intake to differ more fundamentally from the distribution of average intake.

A disadvantage of the Battese procedure is that it requires more days of intake data from each person. The number of days of intake data required from each person depends on the number of moments of the distribution. In general, if the distribution has m moments, then m days of intake data are required from each person. For example, if the distribution is skewed, then it has a third moment, and at least three days of intake data are required from each person

B. A NONPARAMETRIC PROCEDURE PROPOSED BY AICKEN

Aicken developed a fully nonparametric approach for estimating the distribution of usual dietary intake. The procedure is unique in that it does not require a distributional assumption. As with the

NRC adjustment procedure and the Battese procedure, the Aicken procedure is based on the assumptions that (1) the mean value of the error is zero, (2) the error is uncorrelated with usual intake, and (3) the error is uncorrelated with the error on subsequent and previous days. It also assumes that the standard deviation of the error is a linear function of the usual intake.

Aicken tested his procedure with data on the intake of vitamin A by a sample of women from the 1985 CSFII. His procedure produced an estimate of the usual distribution that had a more symmetric shape and a longer right-hand tail than did the NRC adjustment procedure. Unlike the NRC adjustment procedure, which uses only one adjustment factor, the Aicken procedure adjusts each point in the distribution of average dietary intake by a person-specific adjustment factor.

The major disadvantage of the Aicken procedure is that it involves an iterative procedure. Hence, using this procedure to compute the distribution of usual dietary intake may be prohibitively expensive.

APPENDIX C

ESTIMATES OF THE VARIATION IN DAILY DIETARY INTAKE DUE TO INTERVIEW EFFECTS

TABLE C.l

ESTIMATES OF THE VARIATION IN DAILY DIETARY INTAKE CAUSED BY DAY, SEASON, WAVE, AND METHOD OF CONTACT: FULL SAMPLE OF 600 WOMEN

	Dietary Component						
Explanatory Variable	Vitamin C	Calcium	Iron	Protein	Energy	Cholesterol	Saturated Fat
Weekend	3.69 (3.53)	3.37 (17.65)	0.11 (0.25)	1.40 (1.40)	63.41' (32.31)	21.08 (11.14)	0.18 (0.20)
Summer	0.65 (13.8)	-23.53 (76.17)	-0.39 (0.98)	-3.62 (5.47)	-236.08 (125.73)	-41.72 (43.46)	-0.68 (0.79)
Fall	-1.22 (15.23)	-22.57 (76.17)	0.45 (1.09)	-2.25 (6.04)	-181.33 (139.06)	-40.16 (48.07)	-0.50 (0.87)
Winter	-7.72 (8.75)	77.37 (43.77)	0.04 (0.62)	-1.95 (3.47)	59.56 (79.90)	-44.96 (27.62)	1.01 (0.50)
Respondent Interviewed In-Person	-16.94* (4.74)	-83.61' (23.69)	-0.65* (0.34)	-0.25 (l-88)	-54.12 (43.26)	21.00 (14.95)	0.52 (0.27)
Wave 1	11.42 (7.23)	133.69* (36.14)	0.82 (0.52)	2.38 (2.87)	231.61' (65.98)	-21.58 (22.81)	0.01 (0.41)
Wave 2	-6.23 (15.08)	11.23 (73.40)	-0.52 (1.08)	-0.74 (5.98)	247.69 (137.64)	15.45 (47.58)	0.85 (0.86)
Wave 3	-11.76 (15.25)	12.23 (76.25)	-0.91 (1.09)	-0.70 (6.05)	190.98 (139.21)	1.06 (48.12)	0.82 (0.87)
Wave 4	-11.54 (16.19)	33.19 (80.96)	-1.62 (1.15)	-3.64 (6.42)	171.36 (147.80)	-9.96 (51.09)	0.32 (0.93)
Wave 5	3.16 (8.01)	-51.71 (40.04)	0.48 (0.57)	0.48. (3.18)	1.95 (73.10)	12.34 (25.27)	-0.76 (0.46)
R-square	0.009	0.008	0.008	0.005	0.015	0.008	0.006
Sample Size	2,400	2,400	2,400	2,400	2,400	2,400	2,400

SOURCE: Random sample of 600 women taken from the MPR analysis file that contains four days of intake data from the core and low-income samples of the 1985 CSFII.

*Statistically different from zero at the 95 percent confidence level in a two-tail test.

TABLE C.2

ESTIMATES OF VARIATION IN DAILY DIETARY INTAKE CAUSED BY DAY, SEASON, WAVE, AND METHOD OF CONTACT: FULL SAMPLE OF 600 CHILDREN

	Dietary Component						
Explanatory Variables	Vitamin C	Calcium	Iron	Protein	Energy	Cholesterol	Saturated Fat
Weekend	-0.24	-22.51	-0.26	0.01	30.10	5.21	0.19
	(3.21)	(17.44)	(0.29)	(1.03)	(24.00)	(10.21)	(0.18)
Summer	8.94	-60.08	1.10	5.97	212.16*	49.10	-0.83
	(13.46)	(73.10)	(1.21)	(4.33)	(100.59)	(42.78)	(0.76)
Fall	14.67	-18.42	1.83	9.39*	278.00*	62.27	-0.98
	(14.25)	(77.39)	(1.28)	(4.59)	(106.48)	(45.28)	(0.80)
Winter	11.59	20.80	0.36	0.62	116.79*	24.83	-1.18*
	(7.98)	(43.32)	(0.72)	(2.57)	(59.61)	(25.35)	(0.45)
Respondent Interviewed In-Person	-11.26*	-52.15*	0.36	2.22*	-18.34	36.86*	0.64*
	(3.54)	(19.24)	(0.32)	(1.14)	(26.47)	(11.26)	(0.20)
Wave 1	9.93	30.57	0.42	-0.44	48.87*	-16.15	-0.58
	(6.15)	(33.39)	(0.55)	(1.98)	(45.94)	(19.54)	(0.35)
Wave 2	-2.24	-5.33	-1.52	-8.20	-222.61*	-57.80	0.20
	(14.34)	(77.87)	(1.29)	(4.61)	(107.14)	(45.56)	(0.81)
Wave 3	-11.51	6.80	-1.87	-9.62	-266.82.	-74.57	0.73
	(14.27)	(77.51)	(1.28)	(4.59)	(106.65)	(45.35)	(0.81)
Wave 4	-15.35	- 10.74	-1.64	-9.34*	-257.32*	-67.45	0.64
	(14.67)	(79.70)	(1.32)	(4.72)	(109.66)	(46.64)	(0.83)
Wave 5	-8.58	-53.81	-0.46	-1.31	-79.12	-31.26	0.45
	(7.45)	(40.56)	(0.67)	(2.40)	(55.81)	(23.73)	(0.42)
R-square	0.003	0.008	0.006	0.007	0.007	0.008	0.011
Sample Size	2,400	2,400	2,400	2,400	2,400	2,400	2,400

SOURCE: Random sample of 600 children taken from the MPR analysis file that contains four days of intake data from the core and low-income samples of the 1985 CSFII.

*Statistically different from zero at the 95 percent confidence level in a two-tail test.

MATHEMATICA
Policy Research, Inc.

P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

600 Maryland Avenue S.W., Suite 550
Washington, DC 20024-25 12
(202) 484-9220