

# **Assessing the Prevalence of Nutrient Inadequacy**

Alicia L. Carriquiry

*Staff Report 98-SR 87*

November 1998

## **Assessing the Prevalence of Nutrient Inadequacy**

Alicia L. Carriquiry

*Staff Report 98-SR 87*

November 1998

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

*Alicia L. Carriquiry is an associate professor of statistics, Iowa State University.*

This work was partially funded by Contract No. 0009730470 between the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention and the Department of Statistics, Iowa State University.

The author gratefully acknowledges Drs. Wayne Fuller, Helen Jensen, and Sarah Nusser from Iowa State University and Dr. George Beaton from the University of Toronto for their many suggestions and insights. Dr. Wolfgang Kliemann from the Department of Mathematics, Iowa State University, provided valuable insight about some of the material presented in this paper. Kevin Dodd and Gabriel Camaño, graduate research assistants, ISU, produced many of the figures in this paper.

## CONTENTS

List of Figures.....	iv
List of Tables.....	iv
Abstract.....	v
Assessing the Prevalence of Nutrient Inadequacy.....	1
The Joint Distribution Approach to Assessing Nutrient Inadequacy.....	3
The Probability Approach.....	4
The Shortcut Method Proposed by Beaton.....	10
When the Shortcut Method Does Not Work Well.....	13
Intakes and Requirements Are Not Independent.....	13
Requirements Distribution Is Not Symmetric.....	14
The Variance of Requirements Is Larger Than That of Intakes.....	15
Evaluating the Performance of the Shortcut Method.....	16
Estimating Usual Intake Distributions.....	18
References.....	21

## FIGURES

1. Simulated joint distribution of intakes and requirements for 5,000 Individuals. Intakes and requirements are correlated. Along the 45° line, intakes equal requirements .....	3
2. Two simulated risk curves, obtained assuming that requirements for the nutrient are normally distributed with mean 42 and standard deviation 6.5 or 15.0 (dotted). .....	5
3. Simulated risk curve and usual intake distribution. Most of the population is at some risk of nutrient inadequacy. ....	6
4. Simulated risk curve and usual intake distribution. Most of the population is at no risk of nutrient inadequacy. ....	7
5. Simulated risk curve and usual intake distribution. Some individuals in the population are at risk of inadequacy .....	7
6. Simulated joint distribution of intakes and requirements for 5,000 individuals. Usual intake distribution is skewed. The black dot represents at individual in the population.....	11
7. Simulated joint distribution of intakes and requirements for 5,000 individuals. The correlation between $r$ and $i$ is 0.5. ....	14
8. Simulated joint distribution of intakes and requirements for 5,000 individuals. Requirements are assumed to be skewed. ....	15
9. Simulated joint distribution of intakes and requirements for 5,000 individuals. The variance of requirements is large relative to the variance of intakes.....	16
10. Estimated usual intake distribution for vitamin E (in mg of $\alpha$ -TE). The cutoff value is 6 mg of $\alpha$ -TE.....	20

## TABLE

1. Expected number of individuals at risk of nutrient inadequacy (in parentheses) and percentage under- or overestimation of prevalence of shortcut approach. ....	22
--	----

## ABSTRACT

In this study we discuss the problem of estimating the proportion of the population at risk of nutrient deficiency. Ideally, we desire an estimate of the proportion of individuals whose usual nutrient intakes are less than their requirements for the nutrient. This would require knowledge about the *joint distribution* of usual intakes and requirements in the population. In lieu of information about the joint distribution of intakes and requirements, following Beaton<sup>5</sup> we argue that it may suffice to know the mean of the requirement distribution (EAR) together with the distribution of usual intakes in the population. The proportion of individuals with intakes below the EAR for the population is a good approximation to the proportion of individuals with intakes below their requirements, if certain assumptions hold. Those assumptions are as follows: (a) nutrient requirements and usual intakes are independent; (b) the distribution of requirements is symmetric around its mean; and (c) the between-individual variance of requirements is small relative to the between-individual variance in usual intakes.

## ASSESSING THE PREVALENCE OF NUTRIENT INADEQUACY

The problem of assessing the adequacy of nutrient intakes at the population level or, equivalently, of estimating the proportion of individuals at risk of dietary deficiency in a (sub)population has been a long-standing one. Although different analytical approaches to the problem have been proposed, it appears that, at a minimum, we need information on usual nutrient intake distributions as well as on nutrient requirements in the population of interest.

Ideally, the proportion of the population with inadequate intake of a nutrient is estimated from the *joint distribution* of usual intakes and requirements of the nutrient in the population. The joint distribution of intakes and requirements is the distribution of pairs of values  $(r_j, i_j)$  where  $r_j$  and  $i_j$  denote the requirement and the usual intake for the  $j$ th individual in the population, respectively. Here, we define usual intake of a nutrient for an individual as the long-run average intake of the nutrient. Questions such as, What is the probability that a randomly chosen individual has a usual intake below (or above)  $X$  and a requirement below (or above)  $Y$ ? can be answered by looking at the joint distribution of intake and requirement. Similarly, the question of interest, What is the probability that a randomly chosen individual has a *usual intake* that is below his/her requirement? can also be answered using information provided by the joint distribution. Note that the last question could be rephrased to read, What is the proportion of the population at risk of nutrient inadequacy? In contrast, the *univariate distribution* of nutrient requirements  $r$  or of usual intakes  $i$  provides information only about the values of either requirements *or* intakes in the population. The information necessary to estimate the joint distributions of requirements and intakes is very scarce, and thus, the ideal approach is impractical.

Methods to estimate the prevalence of nutrient inadequacy that involve knowledge about the univariate distributions of usual intakes and requirements have been proposed<sup>1</sup> and will be discussed in this study. These methods require that estimates of the usual intake distribution and of the requirements distribution of a nutrient in a (sub)population be combined in some manner, to assess the percentage of the population with inadequate intakes. These approaches involve three steps: (1) estimate the usual intake distribution of the nutrient in the (sub)population; (2) estimate the distribution of requirements of the nutrient in the (sub)population; and (3) combine both distributions to assess the prevalence of inadequate intakes of the nutrient in the (sub)population.

The problem of estimating usual nutrient intake distributions at the population level has been

addressed, and satisfactory solutions have been proposed. In 1986, the National Research Council (NRC) proposed that a measurement error model be used on (perhaps transformed) observed daily intakes, in order to partially remove the effects of day-to-day variability in intakes when estimating usual intake distributions.<sup>1</sup> A group of researchers at Iowa State University (ISU) further refined the NRC approach and developed a statistical procedure for estimating usual nutrient intake distributions that produces reliable estimates with good statistical properties.<sup>2,3,4</sup> Because estimates of usual intake distributions are needed for prevalence assessment, a summary of the features of the method developed at ISU to estimate usual nutrient intake distributions will be presented later in this paper. Application of the ISU method requires multiple 24-hour recalls (two if observations are independent, more if not) on at least a subsample of the individuals.

As has been noted, for example by Beaton<sup>5</sup> and Murphy,<sup>6</sup> we do not, in general, have reliable estimates of the distributions of requirements in the population. Exceptions are a very few nutrients in selected gender-age groups. Often, an estimate of the mean (EAR) and of the variance of the requirement distribution for a nutrient are estimated from observations collected from a small sample of individuals. Letting  $\bar{r}$  and  $S_r$  denote the EAR and the estimated standard deviation of requirements, respectively, and assuming that requirements are *normally* distributed, the Recommended Dietary Allowance (RDA) for a nutrient is computed as

$$RDA = \bar{r} + 1.96 \times S_r .$$

Hence, if requirements for the nutrient are normally distributed in the population, then the proportion of individuals whose requirements exceed the RDA is approximately 2.5%.

The RDA for a nutrient (or some percentage of the RDA) has been used as a cutoff value to assess the prevalence of nutrient inadequacy.<sup>7,8,9</sup> Practitioners estimate the prevalence of nutrient inadequacy by estimating the proportion of the population with intakes below the RDA for the nutrient (and the corresponding gender-age group). This approach leads to overly conservative estimates of prevalence.<sup>1,5,10,11,12</sup> Although at the individual level it is reasonable for clinical dieticians to compare an individual's intake to the RDA, at the population level it is difficult to justify doing so. In this study we argue that Beaton's<sup>5</sup> suggestion that usual intake distributions be compared to the EAR (rather than the RDA) to assess adequacy leads to approximately unbiased estimates of prevalence under certain assumptions.

In the following sections, we first discuss the joint distribution approach to estimating the prevalence of nutrient inadequacy. We then briefly describe what is known as the *probability approach*<sup>1</sup>

to assessing the prevalence of nutrient inadequacy. We then discuss the *shortcut approach* proposed by Beaton<sup>5</sup> and describe the conditions that need to be satisfied for the shortcut method to perform well. While even under ideal conditions the Beaton method leads to prevalence estimates that are biased (albeit slightly), we argue that the loss in accuracy may be offset by the very few assumptions that practitioners must make about the requirements distribution. Illustrative examples of situations in which the shortcut method should *not* be applied are given. Finally, we present the major attributes of the method for estimating usual intake distributions that was developed at ISU in collaboration with the United States Department of Agriculture (USDA).

### The Joint Distribution Approach To Assessing Nutrient Inadequacy

Why would we be interested in estimating the joint distribution of intakes and requirements? Consider the plot presented in Figure 1. Here, we have plotted *simulated* values of requirements and intakes for 5,000 individuals. The pairs  $r_j, i_j$  were drawn from a (bivariate normal) distribution with mean requirement  $\mu_r = 1,200$ , mean intake  $\mu_i = 1,500$ , standard deviation of requirements  $\sigma_r = 200$ , standard deviation of intakes  $\sigma_i = 440$ , and correlation coefficient between requirements and intakes  $\rho_{r,i} = 0.35$ . The bivariate normal distribution was chosen because of its relative simplicity; empirical evidence indicates that the normal distribution is most often *not* a good choice for modeling nutrient intakes.<sup>13,14,15,16</sup>

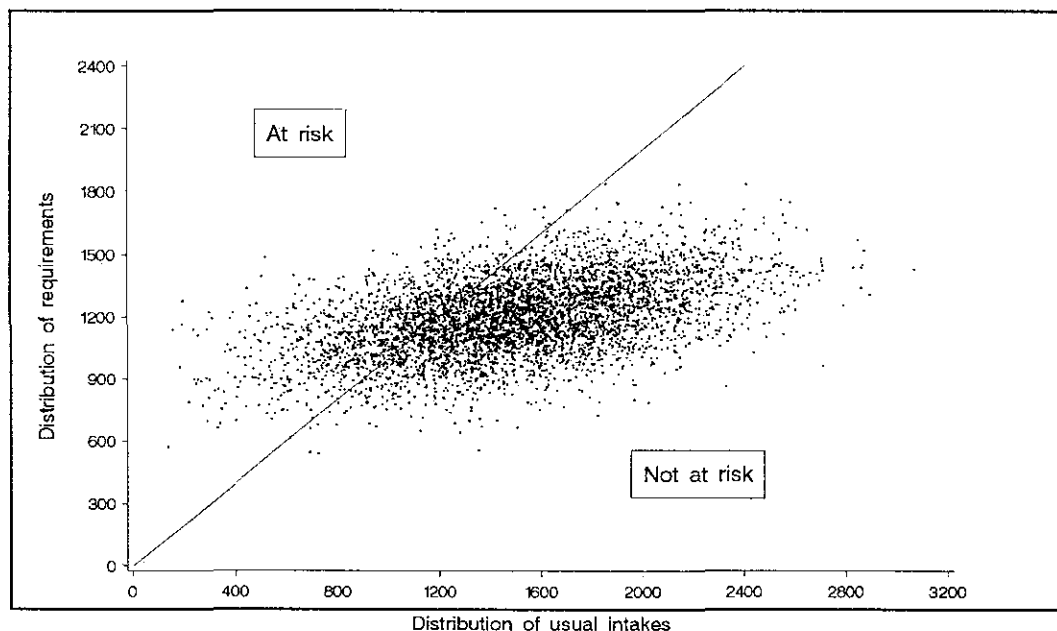


Figure 1. Simulated joint distribution of intakes and requirements for 5,000 individuals. Intakes and requirements are correlated. Along the 45° line, intakes equal requirements.



If we knew the joint distribution of intakes and requirements in a population, we could then estimate the proportion of the population with inadequate intakes by calculating the proportion of individuals whose intakes are below their requirements. In the figure, that proportion would include all individuals above the 45° line, the line where  $r = i$ . For all those individuals, intakes are below their requirements, so that for them,  $i_j - r_j < 0$ .

Formally,

$$\begin{aligned} \Pr(\text{nutrient inadequacy}) &= \Pr(i < r) \\ &= \int_0^{\infty} \int_0^r f_{R,I}(t, s) ds dt, \end{aligned} \quad (1)$$

where  $f_{R,I}(r, i)$  is the joint density of the distribution of intakes and requirements in the population. Given an estimate of the joint distribution  $f_{R,I}$ , obtaining (1) is trivial. The problem is not the actual probability calculation but rather the estimation of the joint distribution of intakes and requirements in the population.

To reduce the data burden for estimating  $f_{R,I}$ , approaches such as the probability approach proposed in the 1986 NRC report<sup>1</sup> or the shortcut approach proposed by Beaton<sup>5</sup> make an implicit assumption: that intakes and requirements are independent random variables. That assumption implies that what an individual consumes of a nutrient is uncorrelated with that individual's requirement for the nutrient. If the assumption of independence holds, then the joint distribution of intakes and requirements can be factorized into the product of the two marginal distributions as follows:

$$f_{R,I}(r, i) = f_R(r) f_I(i), \quad (2)$$

where  $f_R(r)$  and  $f_I(i)$  are the marginal distributions of requirements and of usual intakes of the nutrient, respectively, in the population of interest.

It is important to notice that under the formulation in (2), the problem of assessing the prevalence of nutrient inadequacy becomes tractable. Indeed, as we will discuss in the following sections, methods for reliable estimation of  $f_I(i)$  have been proposed,<sup>2</sup> and data are abundant. Estimating  $f_R(r)$  is still problematic due to scarce requirements data for most nutrients, but the mean and the variance of  $f_R(r)$  can often be computed with some degree of reliability.<sup>5,17,18,19,20,21</sup> In what follows, we discuss approaches for combining  $f_R(r)$  and  $f_I(i)$  for prevalence assessment that require different amounts of information (and assumptions) about the unknown requirements distribution  $f_R(r)$ .

### The Probability Approach

The probability approach to estimating the prevalence of nutrient inadequacy was proposed by the NRC in their 1986 report.<sup>1</sup> The idea is simple. Given a distribution of requirements in the population,

the first step is to compute a *risk curve* that associates intake levels to risk levels under the assumed requirements distribution.

In Figure 2 we show the risk curves for two simulated requirements distributions. Both simulated distributions are normal with mean  $\mu_R = 42$ . The standard deviations of the simulated requirements distributions are 6.5 (solid curve) and 15.0 (dotted curve), respectively.

Formally, the risk curve is obtained from the *cumulative distribution function* (cdf) of requirements. If we let  $F_R(\cdot)$  denote the cdf of the requirements for a nutrient in the population, then

$$F_R(a) = \Pr(\text{requirements} \leq a) .$$

Thus, the cdf  $F_R$  takes on values between 0 and 1. The risk curve  $\rho(\cdot)$  is defined as

$$\rho(a) = 1 - F_R(a) = 1 - \Pr(\text{requirements} \leq a) .$$

In the simulated example in Figure 2,  $F_R(10) \approx 0$  and  $\rho(10) \approx 1$ , while  $F_R(80) \approx 1$  and  $\rho(80) \approx 0$ .

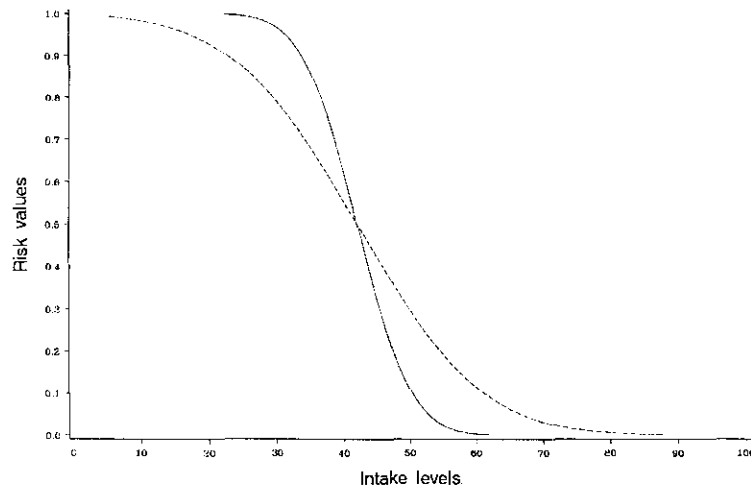


Figure 2. Two simulated risk curves, obtained assuming that requirements for the nutrient are normally distributed with mean 42 and standard deviation 6.5 or 15.0 (dotted).

These risk curves are easy to read. On the  $x$ -axis, the values correspond to intake levels. On the  $y$ -axis, the values correspond to the risk of nutrient inadequacy given a certain intake level. In this

simulated example, the risk of inadequacy for an individual with an intake of 10 is approximately 100% under both simulated risk curves. An individual with an intake of 50 would have a risk of nutrient inadequacy of approximately 12% when the standard deviation of the requirements distribution is 6.5 and a risk of approximately 35% when the standard deviation of the requirements distribution is 15. An individual with an intake of 75 or higher would have a probability of nutrient inadequacy of essentially zero under both simulated risk curves.

How should we combine these risk curves based on assumed requirements distributions with usual intake distributions estimated from dietary survey data?

It seems intuitively appealing to argue as follows. Consider again the simulated risk curves in Figure 2, and suppose we had estimated the usual intake distribution for this simulated nutrient in a population. If that estimated usual intake distribution places very high probability on intake values less than, say, 80, then we would conclude that almost all individuals are at some risk of nutrient inadequacy. If, on the other hand, the usual nutrient intake distribution places most of the probability on intakes *above* 80, then we would be confident that in that population, no one is at risk of nutrient inadequacy.

We illustrate these two cases in the plots in Figures 3 and 4. In Figure 3, the prevalence estimate is approximately 100%, because the distribution of usual intakes in the population says that almost all intakes are between 0 and 80.

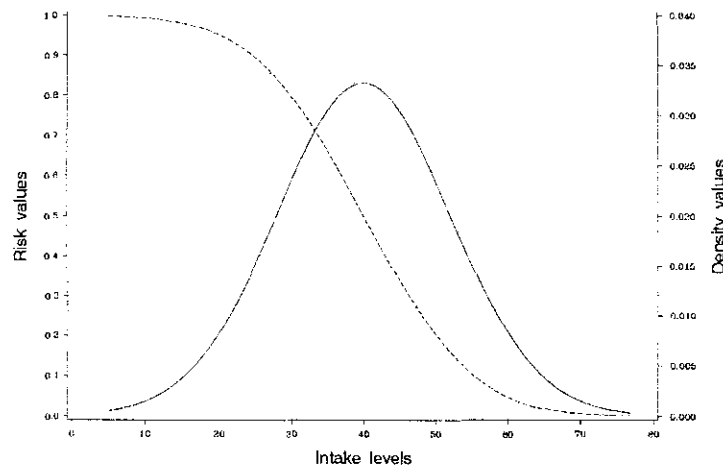


Figure 3. Simulated risk curve and usual intake distribution. Most of the population is at some risk of nutrient inadequacy.

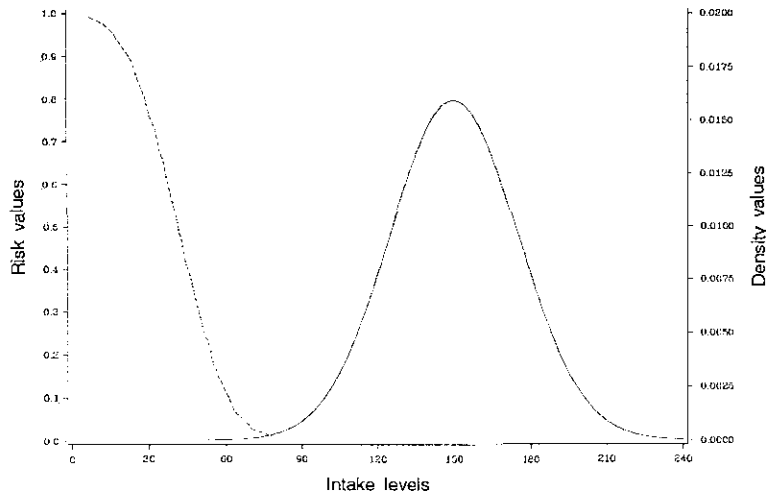


Figure 4. Simulated risk curve and usual intake distribution. Most of the population is at no risk of nutrient inadequacy.

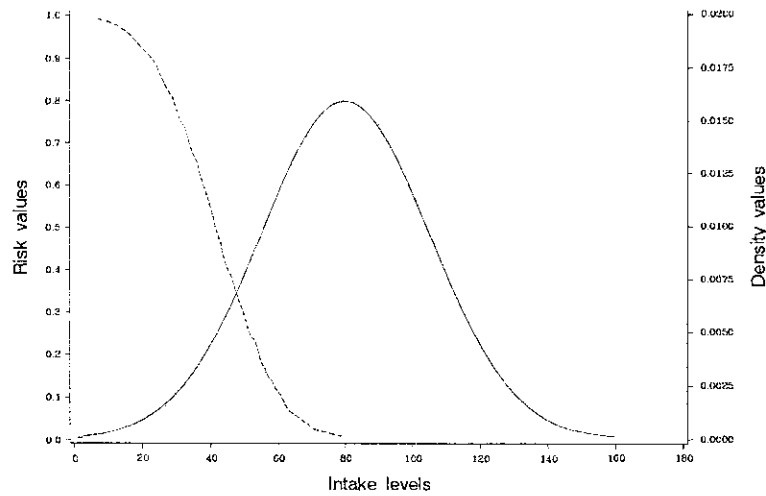


Figure 5. Simulated risk curve and usual intake distribution. Some individuals in the population are at risk of inadequacy

In Figure 4, by contrast, the prevalence estimate would be approximately 0%, because in that population the probability of observing intakes below 80 is very low according to the usual intake distribution.

In general, we would expect that the usual intake distribution and the risk curve for a nutrient show some overlap. A more realistic example is shown in Figure 5. Here we have superimposed a simulated usual intake distribution to the risk curve that arises when requirements are normally distributed with mean 42 and standard deviation 15. The usual intake distribution was simulated as a normal distribution with mean 80 and standard deviation 25.

Notice that in this example, the overlap between the usual intake distribution and the simulated risk curve is small. Indeed, individuals whose intakes are above 70 or so have essentially zero risk of nutrient inadequacy. This was expected in this simulated example, given that the mean intake was set at 80 whereas the mean requirement was set at only 42 for the simulated requirements distribution. Notice also that for an intake level of about 60, the risk of inadequacy is approximately 14%.

The quantity of interest is not the risk associated to a certain intake level but rather the *expected risk of inadequacy* in the population. This expectation (think of it as an average) is taken with respect to the usual intake distribution for the nutrient in the population. In other words, prevalence of nutrient inadequacy is defined as the expected risk given the distribution of intakes in the population. To derive the estimate of prevalence, we first define:

- $y$  denotes intake levels. In Figure 5,  $y$  values are shown between 0 and 160.
- $p(y)$  is the probability, under the usual intake distribution, associated to each intake level  $y$ . In our simulated example,  $p(20)$  and  $p(140)$ , for example, are very small, but  $p(70)$  or  $p(85)$  are very high.
- $\rho(y)$  is the risk calculated from the simulated requirement distribution. For example, for  $y = 20$ , the risk of inadequacy is very high, with  $\rho(20) \approx 1$ , while for  $y = 100$ , the risk is very low, with  $\rho(100) \approx 0$ .

With this notation, the calculation of prevalence is simple. Using the intake levels in the simulated example,

$$\text{Prevalence} = \sum_{y=0}^{\infty} \rho(y)p(y) , \quad (3)$$

where, in practice, the sum is carried out only to intake levels until the risk of inadequacy becomes about 0. In the example in Figure 5, we would add up the product of risk values and intake probabilities only for intake levels between 0 and 80.

Notice that expression (3) is simply a weighted average of risk values, where the weights are given by the probabilities of observing the intakes associated to those risks. Formally, the expected risk is given by

$$\begin{aligned}
 E\{\text{risk}\} &= \int_0^{\infty} \rho(y) dF \\
 &= \int_0^{\infty} \rho(y) f(y) dy,
 \end{aligned}$$

where  $\rho(y)$  denotes the risk value for an intake level  $y$ ,  $F$  is the usual intake distribution, and  $f(y)$  is the value of the usual intake density at intake level  $y$ .

The NRC report<sup>1</sup> suggests a simpler computational method to obtain an estimate of prevalence using the probability approach. Rather than adding up many products  $\rho(y) p(y)$  associated to different values of intakes, one can group intakes by constructing  $m$  “bins”. In the example, we could group all intakes between 0 and 10 into a bin, all intakes between 20 and 30 into another bin, and so forth. The probabilities associated to each bin are simply the frequencies of intakes in the population that “fall into” each bin. (These frequencies are determined by the usual intake distribution in the population.) The average risk associated to intakes in a bin is approximated as the risk associated to the middle element in the bin. In the example, all intakes between 50 and 60 would have a risk of approximately 18% under the risk curve. To compute prevalence using this grouped version, simply use expression (3) but add over the  $m$  bins using the values of  $\rho$  associated to the midpoint of each bin and  $p(\cdot)$  of the bin. An example of this computation is given on page 28, Table 5-1, of the NRC report.<sup>1</sup>

The probability approach requires the following information:

- Intakes and requirements are assumed to be independent. Under this independence assumption, the joint distribution of intakes and requirements can be factorized, as we discussed in a previous section, into the corresponding marginal distributions.
- The mean and the variance of the requirement distribution must be known.
- Practitioners must know the *form* of the requirements distribution (normal, lognormal, etc.) or be willing to *assume* a form for the requirement distribution. Without a form for the distribution, risk curves such as those shown in Figures 2 and 5 cannot be computed.

In general, researchers assume that requirements distributions are normal, with mean and variance as estimated from experimental data. Even under normality, however, an error in the estimation of either the mean or the variance (or both) of the requirements distribution may lead to biased prevalence estimates. The NRC report<sup>1</sup> provides various examples of the effect of changing the mean and the variance of the requirement distribution on prevalence estimates. What the NRC report<sup>1</sup> does not mention is that an incorrect specification of the form of the requirements distribution may lead to even more elaborate biases.

### The Shortcut Method Proposed By Beaton

The probability approach described in the previous section is simple to apply but requires that practitioners choose a form for the requirements distribution. For certain nutrients such as iron, the assumption of normality is inappropriate, and for all nutrients, data are not abundant enough to check whether the normality assumption is reasonable. Thus, a method that does not require such strong parametric assumptions would be very appealing. The shortcut method proposed by Beaton<sup>5</sup> requires no such assumptions and, under certain conditions, should be preferred to the probability approach.

The shortcut method is very simple: estimate prevalence of inadequate intakes as the proportion of the population with unusual intakes below the average requirement (EAR). In the simulated example in the previous section, prevalence would be estimated as the proportion of individuals in the population with usual intakes below 42.

To understand how the shortcut method works, we go back to the section where we discussed the joint distribution approach to estimating prevalence. Refer to Figure 6, which shows a simulated joint distribution of usual intake and requirement. To generate the joint distribution, we simulated usual intakes and requirements for 5,000 individuals, from independent  $\chi^2$  distribution with seven degrees of freedom and a normal distribution, respectively. The usual intake distribution was rescaled to have mean 1,500 and standard deviation 400. The normal distribution used to represent requirements had mean 1,200 and standard deviation 200. Note that intakes and requirements are uncorrelated and that the usual intake distribution is skewed. An individual whose intake is below the *mean requirement* is not necessarily at risk. For example, the individual marked with a black dot in Figure 6 has an intake below the mean requirement EAR but is not at risk because for that individual,  $i - r > 0$ .

Because we base our inferences on the joint, rather than the univariate, distributions, we may well find that an individual consuming a nutrient at a level below the mean of the population requirement may be satisfying his/her own requirements and thus may not be at risk. That is the case with all the individuals in the plot that appear below the 45° line and to the left of the  $i = 1,200$  line, in the triangle labeled B.

We have agreed that in order to estimate prevalence we would need to “count” those individuals in Figure 6 who appear *above* the 45° line, because for them,  $i < r$ . We have also agreed on the fact that this is not a practical method, since we do not have information needed for estimating the joint distribution. Can we approximate this proportion in some other way? The probability approach in the previous section is one such approximation. The shortcut method of Beaton<sup>5</sup> is another such approximation, and we now describe its fundamentals.

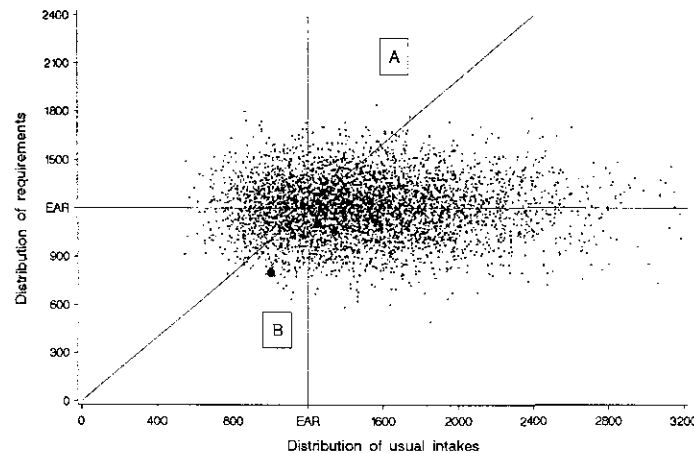


Figure 6. Simulated joint distribution of intakes and requirements for 5,000 individuals. Usual intake distribution is skewed. The black dot represents an individual in the population.

Refer to the plot in Figure 6. Notice that the number of individuals with intakes to the *left* of the vertical EAR=1,200 reference line is more or less the same as the number of individuals *over* the 45° line. That is, it appears from the figure that counting the proportion of individuals with intakes below the EAR yields approximately the same number as counting the number of individuals with intakes below their own requirements. In other words, the number of individuals in the triangle labeled A appears to be approximately the same as the number in the triangle labeled B in Figure 6. If this is the case, we can then assess prevalence of inadequate intakes as long as we have an estimate of the usual nutrient intake distribution (which is almost always available) and of the mean requirement in the population, or EAR, which can be obtained reliably from relatively small experiments.

When can we justify estimating the proportion of the population at risk via the proportion of the population with intakes below the mean requirement? This question is best answered from a statistical viewpoint. Inferences about dietary risk can be based on intake distributions and mean requirements, *if* we are willing to make the following assumptions: (1) intakes and requirements are independent; in statistical terms, this means that their joint distribution can be estimated as the product of the marginal distributions, as in expression (2); (2) the distribution of requirements is symmetric (not necessarily normal); and (3) the variance of the distribution of requirements is small relative to the variance of the distribution of intakes. That is, there is more between individual variation in intakes than in requirements in the population.



A few comments about these conditions are in order. First, these conditions are rather mild, in particular when compared with the assumptions that are implicit in the probability method described previously. Notice that condition 1 is implicit in the probability method. Condition 2 is satisfied for most nutrients; furthermore, assuming symmetry of the requirements distribution is much less demanding than assuming that the form of the distribution is known, as is required by the probability approach. One exception to this symmetry rule is iron, for which requirements in women are known to exhibit skewness. Finally, condition 3 is satisfied in general, a possible exception being institutionalized populations where intakes may not vary significantly from one individual to the next. We will discuss the impact on prevalence estimates when any of these three conditions are not satisfied later in this report. Finally, notice that *no* conditions are imposed on the form of the usual intake distribution, except that its variance is large relative to the variance of the requirement distribution. The simulation population shown in Figure 6 is one for which the three conditions are satisfied.

If we were interested in estimating the proportion of the population at risk of dietary deficiency from data satisfying the assumptions above, we would not require knowledge of the joint distribution of intakes and requirements. Rather, and as was pointed out by Beaton,<sup>5</sup> we would only need reliable estimates of (1) the mean of the requirement distribution for the population, and (2) the distribution of usual intakes in the population. In particular, we would need good estimates of tail probabilities of the intake distribution, such as percentiles as low as the second or third.

Can we justify this approximation? Notice from the plot in Figure 6 that assessing the proportion of the population at risk is equivalent to calculating the proportion of individuals with intake below their requirement, as before. That is, we need the proportion of individuals above the 45° line. Let that proportion of individuals at risk be denoted by  $p_r$ . If the assumptions listed previously are met (i.e., if our data, were they available, were to “look” like the data in Figure 6), then we would obtain a good approximation to  $p_r$  by calculating instead the proportion of individuals with intake below the mean requirement, or equivalently, to the left of the vertical reference line in the plot. Denote the proportion of individuals with intake below the mean requirement by  $p_{\bar{r}}$ . Note that under the previous assumptions,  $p_{\bar{r}}$  is a good approximation to  $p_r$ : those individuals in the top triangle labeled A in the plot, that should be counted but are not, are instead compensated by those individuals in the bottom triangle labeled B in the plot, that should not be counted but are. In other words, under the assumptions, the proportion of individuals with intake below the mean requirement but above their own requirement is approximately the same as the proportion of individuals with intake above the mean requirement but still below their individual requirement. Thus, the proportion of individuals above the 45° line can be approximated by

the proportion to the left of the  $i = \bar{r}$  (intake = average requirement) line.

In the simulated example shown in Figure 6 the true number of individuals with usual intake below their requirement is 1,405. The number that would be estimated using the shortcut approach of Beaton<sup>5</sup> is 1,350. The underestimation of prevalence in this example is only approximately 1%.

We conducted a small Monte Carlo experiment to assess the performance of the shortcut approach under varying assumptions for the distributions of usual intakes and of requirements. The results from the simulation experiment are given in the section entitled Evaluating the Performance of the Shortcut Method.

### When the Shortcut Method Does Not Work Well

The shortcut method for estimating prevalence provides a good approximation to the true prevalence in the population when the three conditions listed in the previous section are met. To better understand the consequences of violating any of the three assumptions, we have simulated data sets that represent the joint distribution of intakes and requirements for nutrients such as iron, niacin, or energy, in which one or more of the conditions may be violated. We discuss departures from each of the assumptions separately.

#### Intakes and Requirements Are Not Independent

What would happen if the assumption of independence of intakes and requirements does not hold? This is the case, for example, of energy and niacin,<sup>1,5,18</sup> in which the amount of the nutrient consumed is correlated to the requirement of the nutrient by the individual.

In this case, the approximation that consists of estimating the proportion of individuals with intakes below the mean population requirement would not be a good one. In Figure 7 we show a simulated joint distribution of intakes and requirements obtained by generating 5,000 pairs  $(r,i)$  from a bivariate normal distribution with mean requirement equal to 1,200 and mean intake equal to 1,500. Here, the correlation between intakes and requirements is relatively high, 0.50.

Notice that in this case we cannot approximate the number of points above the 45° line by counting the points to the left of the vertical  $i = 1200$  line. The number of points in the lower triangle labeled B is larger than the number of points in the upper triangle labeled A. In this case, using the shortcut method proposed by Beaton<sup>5</sup> would lead to a significant *overestimation* of the prevalence of inadequate nutrient intakes in the population. In this simulated example, the true prevalence is 21%, whereas the estimated prevalence using the shortcut method is approximately 27%. This means that in this population of 5,000 individuals we would be overestimating the number at risk of nutrient inadequacy by approximately 300 individuals.

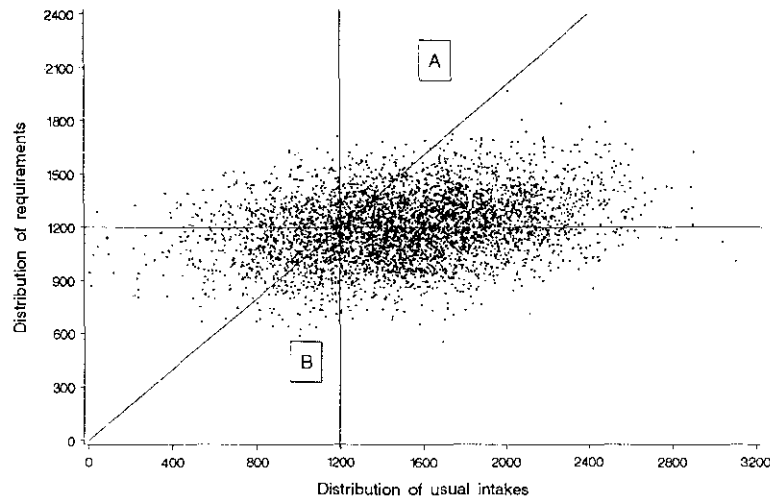


Figure 7. Simulated joint distribution of intakes and requirements for 5,000 individuals. The correlation between  $r$  and  $i$  is 0.5.

### Requirements Distribution Is Not Symmetric

The assumption of a symmetric requirements distribution is untenable for some nutrients in some (sub)populations. For example, it is known that menstruating women require increased amounts of iron,<sup>1,19</sup> and thus it is appropriate to model requirements using a skewed distribution with a long tail to the right.

When the assumption of symmetry of the requirements distribution is not satisfied, the shortcut approach of Beaton<sup>5</sup> produces biased prevalence estimates.

Refer to Figure 8 where we show 5,000 simulated pairs of intakes and requirements. Here, skewed ( $\chi^2$ ) distributions were used to represent both the intakes and the requirements distributions. Notice that values spread farther in the “north” and “east” directions than in the “south” and “west” directions.

In this example, the number of individuals to the left of the intake = EAR line is *less* than the number of individuals above the 45° line. Therefore the number of individuals with intakes below the EAR is *not* a good approximation to the number of individuals with intakes below their own requirements. Thus, the shortcut approach produces an underestimation of the true prevalence of inadequate intakes. The extent of the underestimation in this simulated example is approximately 4% of the population, or about 200 individuals in the population of 5,000.

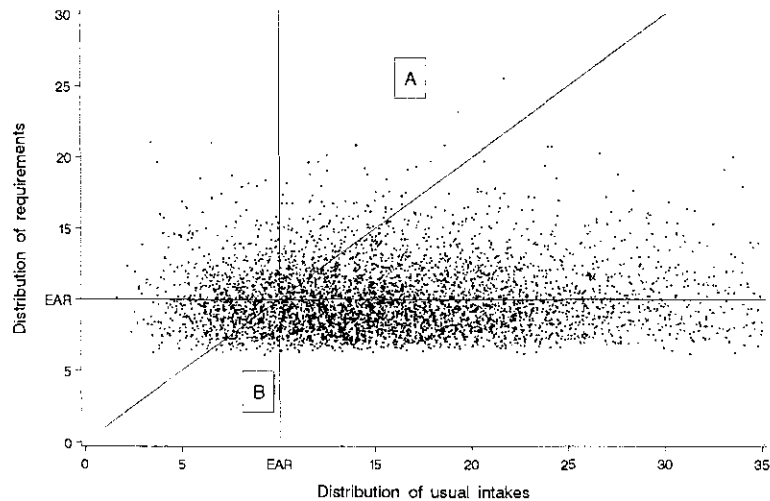


Figure 8. Simulated joint distribution of intakes and requirements for 5,000 individual. Requirements are assumed to be skewed.

### The Variance of Requirements Is Larger Than That of Intakes

Finally, the third condition that needs to be satisfied is that the variance of the requirements distribution must be relatively small compared to the variance of the intake distribution.

Figure 9 shows 5,000 simulated pairs of intakes and requirements that might be representative, for example, of an institutionalized population. Notice that in this simulated population the mean intake is higher than the mean requirement, and the variance of intakes is significantly smaller than the variance of requirements.

Again, the shortcut method produces an estimate of prevalence that is noticeably biased relative to what would have been estimated had the joint distribution been available. In the example shown in Figure 9, the shortcut method produces a seriously biased prevalence estimate. Whereas the true number of individuals at risk in the population of 5,000 is 1,414, the shortcut estimate is only 481 individuals.

Although the figures we have presented show specific illustrative examples, it is important to note that whether the shortcut method produces an overestimate or an underestimate of the true prevalence in the population depends on other factors such as the value of the mean intake relative to the mean requirement. For example, if in the plot shown in Figure 9 we had set the mean intake to be equal to the mean requirement, then of course the bias of the shortcut approach would have been essentially zero,

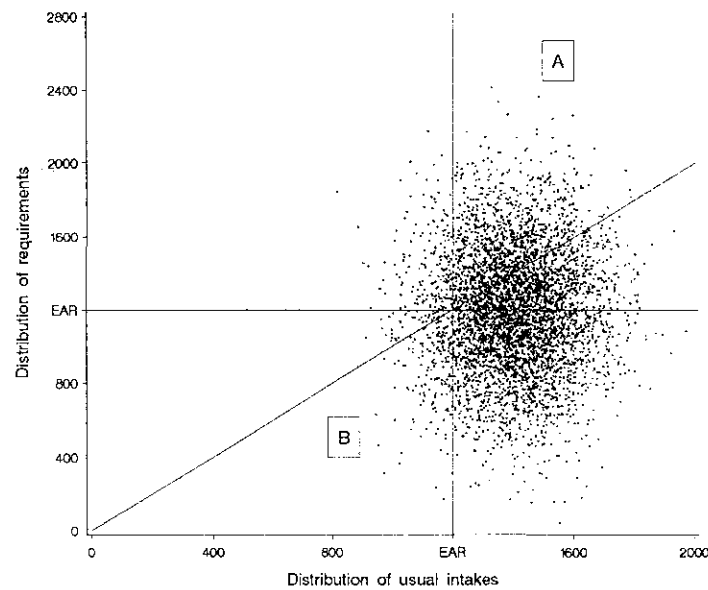


Figure 9. Simulated joint distribution of intakes and requirements for 5,000 individuals. The variance of requirements is large relative to the variance of intakes.

even though condition 3 is violated. The effects of the relative sizes of the mean intake and EAR, as well as of the standard deviation of intakes and requirements, are highlighted in the Monte Carlo study presented in the next section.

### Evaluating the Performance of the Shortcut Method

How well does the shortcut method perform? We have argued that, as long as the three conditions hold, the method proposed by Beaton<sup>5</sup> to assess the prevalence of nutrient inadequacy works well.

In this section, we present the results of a Monte Carlo study conducted to assess the performance of the shortcut method under 28 different scenarios constructed as follows:

- The EAR was kept constant at a value of 1,200.
- The standard deviation  $\sigma_i$  of the simulated usual intake distribution was kept constant at 400.
- The mean of the usual intake distribution  $\mu_i$  was set at seven different values: 70%, 80%, 90%, 100%, 110%, 120%, and 130% of the EAR.
- The standard deviation of the requirements distribution  $\sigma_r$  was set at four different values: 45%, 85%, 145%, and 185% of  $\sigma_i$ .

In all 28 cases, the requirements distributions were simulated to be normal with mean EAR and variance  $\sigma_r^2$ , and the usual intake distributions were simulated as  $\chi^2$  distributions with 10 degrees of

freedom, rescaled to have mean  $\mu_i$  and variance  $\sigma_i^2$ . Intakes and requirements were simulated to be independent. That is, conditions 1 and 2 hold in all 28 scenarios.

As in the previous examples, we generated populations of size 5,000. For each of the 28 combinations of  $\mu_i$ , EAR,  $\sigma_i$ , and  $\sigma_r$  we simulated 3,000 replications of such populations.

Results from the experiment are given in Table 1. We present the *expected* number of individuals at risk of nutritional deficiency in parentheses. This expectation is computed as the average number, over the 3,000 replications, of individuals whose intakes were below their requirements. For example, when the mean of the usual intake distribution is 70% of the EAR and the standard deviation of requirements is 45% of the standard deviation of intakes, the expected number of individuals at risk is 4,261 (out of 5,000).

Table 1. Expected number of individuals at risk of nutrient inadequacy (in parentheses) and percentage under- or overestimation of prevalence of shortcut approach.

$\mu_i$	$0.45 \sigma_i$	$0.85 \sigma_i$	$1.45 \sigma_i$	$1.85 \sigma_i$
0.7 EAR	+1.7% (4,261)	+5.6% (4,066)	+12.4% (3,727)	+16.1% (3,542)
0.8 EAR	+2.0% (3,877)	+6.1% (3,671)	+12.1% (3,370)	+15.0% (3,226)
0.9 EAR	+2.0% (3,362)	+5.6% (3,184)	+9.8% (2,975)	+11.6% (2,884)
EAR	+1.5% (2,724)	+3.4% (2,627)	+4.9% (2,552)	+5.4% (2,530)
1.10 EAR	+0.1% (2,007)	-0.5% (2,040)	-2.2% (2,124)	-3.2% (2,173)
1.20 EAR	-1.8% (1,298)	-5.3% (1,471)	-10.0% (1,707)	-12.3% (1,825)
1.30 EAR	-3.4% (705)	-8.8% (976)	-15.8% (1,323)	-19.3% (1,497)

The expected over- (or under) estimation of the proportion at risk that results from using the shortcut approach in each of the 28 cases is given as a percentage of the expected proportion at risk. For example, in the first cell of the table the estimated number of individuals at risk had a mean (over the 3,000 replications) of 4,333, or 1.7% higher than the “truth”.

By moving from row to row in Table 1 (along a column) we can assess the effect of the relative sizes of  $\mu_i$  and EAR. Similarly, moving across columns (within a row) we can see the effect of the relative sizes of  $\sigma_r$  and  $\sigma_i$  on the performance of the shortcut approach for estimating prevalence.

We note the following:

- The shortcut approach appears to perform well in most situations as long as  $\sigma_r < \sigma_i$ . That is, when condition 3 of Beaton's<sup>5</sup> shortcut method is satisfied (as in the first two columns of the table) then the amount of over- or underestimation that results from applying the shortcut approach is small.
- When  $\sigma_i > \sigma_r$ , the biases that result from application of the shortcut approach are unacceptably high. For example, the fourth column in Table 1 suggests that we would be underestimating the true proportion at risk by almost 20% when the mean usual intake is high relative to the EAR. A similar scenario was illustrated in Figure 9.
- As expected from statistical theory, the biases approach zero as the mean usual intake approaches the EAR. Refer to rows three and four in Table 1. This is true even when condition 3 does not hold. In fact, even though we do not show it in the simulation results, the bias of the shortcut approach will be small *regardless* of whether conditions 1 *and* 3 hold or not, as long as  $\mu_i \approx \text{EAR}$ . This is a strong observation that suggests that the shortcut approach proposed by Beaton<sup>5</sup> will adequately approximate prevalence when mean usual intake and mean requirements are similar, for many possible forms of the (unknown) joint distribution of requirements and usual intakes.

Results from the Monte Carlo study shown in Table 1 are encouraging. The shortcut approach for estimating prevalence of inadequate nutrient intake at the population level performs well in many situations. Indeed, it appears to be robust to departures from the conditions listed in the section on the shortcut method.

### Estimating Usual Intake Distributions

It is apparent from the preceding discussion that reliable estimates of usual intake distributions are needed to assess the proportion of the population at risk of nutrient inadequacy. Estimating the distribution of usual intakes of nutrients from dietary intake data is a difficult problem. Several characteristics of dietary intake data need to be addressed in the estimation procedure. Some of these characteristics include:

- Within-individual variation in intakes. Sometimes the within-individual variation is very large relative to the between-individual variance.<sup>2,5,22,23</sup>
- Skewed (not normal nor symmetric) intake distributions. In some nutrients, there appears to be a very

long tail to the right of the distribution. This skewness is more pronounced when supplements are included as part of the diet.<sup>2,15</sup>

- Heterogeneous within-individual variances. Evidence indicates that as individual intakes increase, so do within-individual variances. Further, the relationship between an individual's mean and his/her variance is often not linear.<sup>2,3,16</sup>
- Day-to-day correlation in intakes when dietary data are collected over consecutive days.<sup>24</sup>
- Complex survey designs that require the use of sampling weights when drawing population inferences, such as estimation of the percentage of the population with intakes below (or above) a given cutoff.

The method developed at ISU for estimating usual intake distributions addresses all these attributes of dietary data. Data required for the implementation of the procedure include two or more independent, 24-hour recalls from at least some of the individuals in the sample. If dietary data are collected over consecutive days, then three or more 24-hour recalls should be obtained from at least some individuals in the sample. The method was developed by expanding on the procedure proposed in the NRC report.<sup>1</sup> It is based on the same measurement error model that states that the intake we observe for any individual on any day is equal to the sum of that individual's true usual intake and a measurement error for that individual and that day. Details about the procedure are given in the nutrition<sup>3</sup> and the statistics literature.<sup>2,4,16</sup>

The method developed at ISU<sup>2</sup> produces estimates of percentiles that are less biased than those produced by other methods. The bias reduction can be very large, even relative to the method proposed in the NRC report.<sup>1</sup> The performance of the different methods available was tested by extensive simulation experiments.<sup>2</sup> In all cases, the ISU method produced better (in the minimum mean squared error sense) estimates of percentiles than any other procedure. This was more noticeable in the tails of the distribution, where estimation is the most difficult.

As an illustration, we present estimates of the usual intake distribution of vitamin E from intake data collected in the 1989-91 Continuing Survey of Food Intakes by Individuals (CSFII) from 887 men aged 60 years and older. Estimated distributions are presented in Figure 10.

The three curves in the plot correspond to estimates of the usual intake distribution obtained by the following three procedures: (1) compute the means of three days and fit a distribution to the observed individual means; (2) apply the method proposed in the NRC report;<sup>1</sup> and (3) apply the method developed at ISU.<sup>2,3</sup>

It is clear from the plot in Figure 10 that the estimates of the proportion of the population with intakes below the (arbitrary) cutoff value of 6 mg  $\alpha$ -TE vary greatly depending on the method used to



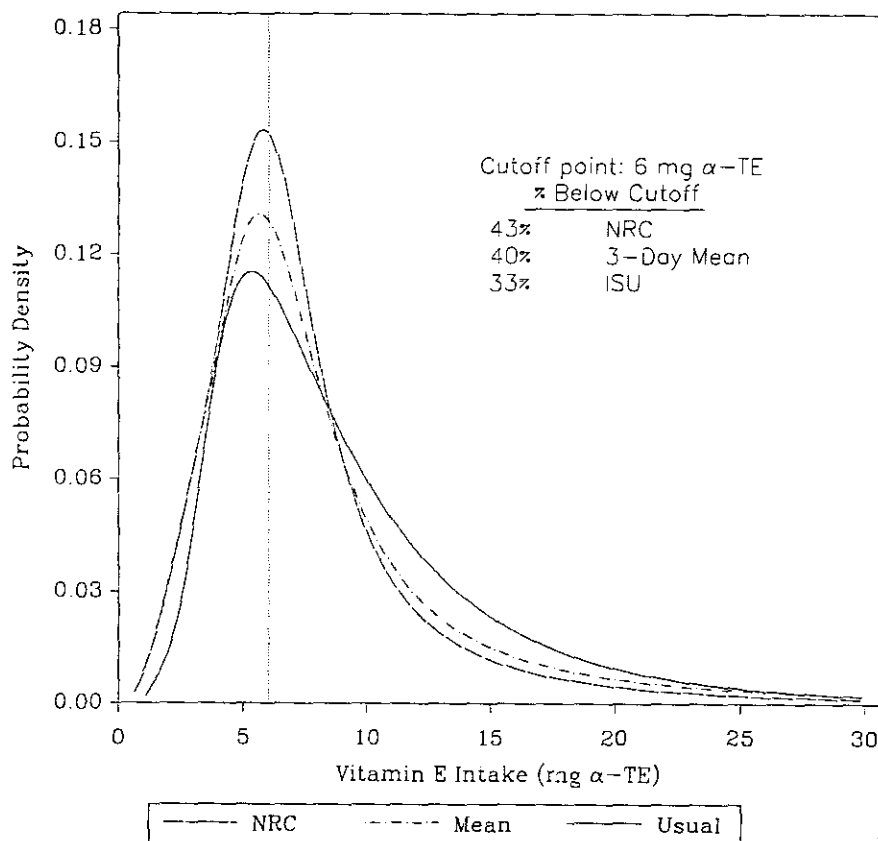


Figure 10. Estimated usual intake distribution for vitamin E (in mg of  $\alpha$ -TE). The cutoff value is 6 mg of  $\alpha$ -TE

estimate the distribution. Estimated values for the proportion of the population with intake below the cutoff is almost 10% lower in the ISU method than in the NRC procedure. In addition, we estimated the proportion of the population below the cutoff when only the first day of data is used for each individual in the sample. The estimated proportion below the cutoff was 45%, significantly higher than the estimate obtained from the ISU method. This result was to be expected, since estimating usual intake distributions from only one day of data does not take into account the within-individual variation in intakes. The large differences among the four estimates highlights the importance of using the appropriate methods for estimating usual intake distributions.

## REFERENCES

1. National Research Council. *Nutrient Adequacy: Assessment Using Food Consumption Surveys*. Report of the Food and Nutrition Board. Washington, D.C., National Academy of Sciences. 1986.
2. Nusser, S. M., Carriquiry, A. L., Dodd, K. W., Fuller, W. A. A semiparametric transformation approach to estimating usual daily intake distributions. *J. Am. Stat. Assoc.* 1996; **91**:1440 - 1449.
3. Guenther, P. M., Kott, P. S., Carriquiry, A. L. Development of an approach for estimating usual nutrient intake distributions at the population level. *J. of Nutr.* 1997; **127**:1106 - 1112.
4. Carriquiry, A. L., Nusser, S. M. Analyzing dietary intake data. Iowa State University, Department of Statistics, preprint. Ames, IA 50011. 1998.
5. Beaton, G. H. Criteria of an adequate diet. In: Shils, R. E., Olson, J. A., Shike, M., eds. *Modern Nutrition in Health and Disease*. Lea and Febiger, Philadelphia 1994; 1491 - 1505.
6. Murphy, S. P. The need for different recommendations for individual and population intakes. Unpublished manuscript, University of California, Berkeley, 1994.
7. Lörstad, M. H. Recommended intake and its relation to nutrient deficiency. *FAO Nutrition Newsletter*. 1971; **9**:18 - 31.
8. Peterkin, B. B., Kerr, R. L., Hama, M. Y. Nutritional adequacy of diets of low-income households. *J. Nutr. Educ.* 1982; **14**:102 - 104.
9. Life Sciences Research Office. *Guidelines for use of dietary intake data*. Anderson, S. A. ed. Federation of American Societies for Experimental Biology, Bethesda, MD. 1986.
10. Hegsted, D. M. Problems in the use and interpretation of the Recommended Dietary Allowances. *Ecol. Food Nutr.* 1972; **1**:255 - 265.
11. Beaton, G. H. What do we think we are estimating? In: Beal, V. A., Laus, M. J., eds. *Proceedings of the Symposium on Dietary Data Collection, Analysis, and Significance*. Res. Bull. No. 675, University of Massachusetts at Amherst. 1982.
12. Beaton, G. H. *Empiric relationships between intake and requirement distributions*. Report submitted to the Food and Nutrition Board, Institutes of Medicine. 1995.
13. Sempos, C. T., Johnson, N. E., Smith, E. L., Gilligan, C. Effects of intraindividual and interindividual variation in repeated dietary records. *Am. J. Epidem.* 1985; **121**:120 - 130.
14. Aickin, M., Rittenbaugh, C. Woolf, K. A general method for adjusting dietary data to obtain stable estimates of usual intake. *Am. J. of Clin. Nutr.* 1994; **59**(suppl.):303S (abstr.).

15. Emrich, J. E., Dennison, D., Dennison, K. Distributional shape of nutrition data. *J. of Am. Diet. Assoc.* 1989; **89**:665 - 670.
16. Carriquiry, A. L., Dodd, K. W., Nusser, S. M. *Estimating adjusted intake and blood biochemical measurement distributions for NHANES III*. Report submitted to the National Center for Health Statistics, Centers for Disease Control and Prevention. Department of Statistics, Iowa State University. 1997.
17. Beaton, G. H., Chery, A. Protein requirements of infants: a reexamination of concepts and approaches. *Am. J. Clin. Nutr.* 1988; **105**:914 - 923.
18. FAO/WHO/UNU. Energy and Protein Requirements. Report of a Joint FAO/WHO/UNU Expert Consultation. *WHO Tech. Rep. Ser.* No. 724. 1985.
19. FAO/WHO. Requirements of Vitamin A, Iron, Folate, and Vitamin B12. Report of a Joint FAO/WHO Expert Consultation. *FAO Food and Nutrition Ser.* No. 23. 1988.
20. Beaton, G. H. Recommended dietary intakes: individuals and populations. In: *Modern Nutrition in Health and Disease*, 9th ed. Williams, G., Wilkins, S., eds. Lea and Febiger, Philadelphia 1998.
21. Dewey, K. G., Beaton, G. H., Fjeld, C., Lonnerdal, B., Reeds, P. Protein requirements of infants and children. *Eur. J. Clin. Nutr.* 1996; **50**(suppl.):S119 - S150.
22. Borrelli, R., Cole, T. J., Di Biase, G., Contaldo, F. Some statistical considerations on dietary assessment methods. *Eur. J. Clin. Nutr.* 1989; **43**:453 - 463.
23. Beaton, G. H., Burema, J., Rittenbaugh, C. Errors in the interpretation of dietary assessments. *Am. J. Clin. Nutr.* 1997; **65**(suppl.):1100S - 1107S.
24. Carriquiry, A. L., Fuller, W. A., Goyeneche, J. J., Jensen, H. H. Estimated correlations among days for the combined 1989-91 CSFII. *Dietary Assessment Research Series 4*. CARD Staff Report 95-SR77. Center for Agricultural and Rural Development, Iowa State University, Ames. 1995.