

Assessing the prevalence of nutrient inadequacy

Alicia L Carriquiry*

Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, USA

Submitted 26 May 1998: Accepted 16 September 1998

Abstract

Objective: To describe an approach for assessing the prevalence of nutrient inadequacy in a group, using daily intake data and the new Estimated Average Requirement (EAR).

Design: Observing the proportion of individuals in a group whose usual intake of a nutrient is below their requirement for the nutrient is not possible in general. We argue that this proportion can be well approximated in many cases by counting, instead, the number of individuals in the group whose intakes are below the EAR for the nutrient.

Setting: This is a methodological paper, and thus emphasis is not on analysing specific data sets. For illustration of one of the statistical methods presented herein, we have used the 1989–91 Continuing Survey on Food Intakes by Individuals.

Results: We show that the EAR and a reliable estimate of the usual intake distribution in the group of interest can be used to assess the proportion of individuals in the group whose usual intakes are not meeting their requirements. This approach, while simple, does not perform well in every case. For example, it cannot be used on energy, since intakes and requirements for energy are highly correlated. Similarly, iron in menstruating women presents some difficulties, due to the fact that the distribution of iron requirements in this group is known to be skewed.

Conclusions: The apparently intractable problem of assessing the proportion of individuals in a group whose usual intakes of a nutrient are not meeting their requirements can be solved by comparing usual intakes to the EAR for the nutrient, as long as some conditions are met. These are: (1) intakes and requirements for the nutrient must be independent, (2) the distribution of requirements must be approximately symmetric around its mean, the EAR, and (3) the variance of the distribution of requirements should be smaller than the variance of the usual intake distribution.

Keywords
EAR
Inadequate intakes
Requirements
Probability approach
Intake distribution

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. While different analytical approaches to the problem have been proposed, it appears that, at a minimum, we need information on usual nutrient intake distributions, and on nutrient requirements, in the population of interest.

Ideally, the proportion of the population with inadequate intakes 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* distributions of nutrient requirements r or of usual intakes i provide information only about the values of either requirements or intakes in the population. The information necessary to estimate the *joint* distribution 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 proposed¹ and will be discussed in this paper. These approaches require that estimates of the usual intake distribution and the requirements distribution of a nutrient in a subpopulation be combined in some manner, to assess the percentage

of the population with inadequate intakes. These methods then, 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.
- 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, to partially remove the effects of day-to-day variability in intakes when estimating usual intake distributions¹. 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 properties²⁻⁴. 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 for estimating the distribution of usual nutrient intakes requires multiple 24-h recalls (two if observations are independent, more if not) on at least a subsample of the individuals.

As has been noted, for example, by Beaton⁵ and Murphy⁶, we do not, in general, have reliable estimates of the distributions of requirements in the population. Exceptions are very few nutrients in selected gender-age groups. Often, an estimate of the mean (EAR) and 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, 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 cut-off value to assess the prevalence of nutrient inadequacy⁷⁻⁹. 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^{1,5,10-12}. While at the individual level it is reasonable for clinical dietitians to compare an individual's intake to the RDA, at the population level it is difficult to justify doing so. In this paper, we argue that Beaton's⁵ 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 remainder, 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*¹ to assess prevalence of nutrient inadequacy. We then discuss the EAR cut-point *approach* proposed by Beaton⁵ and describe the conditions that need to be satisfied for the EAR cut-point method to perform well. While even under ideal conditions the cut-point method leads to prevalence estimates that are biased (albeit slightly so), we argue that the loss in accuracy may be offset by the very few assumptions about the requirements distribution that practitioners must make. Illustrative examples of situations in which the cut-point method should *not* be applied are given. Finally, we present the major attributes of the ISU method for estimating usual intake distributions, that was developed at ISU in collaboration with the US 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

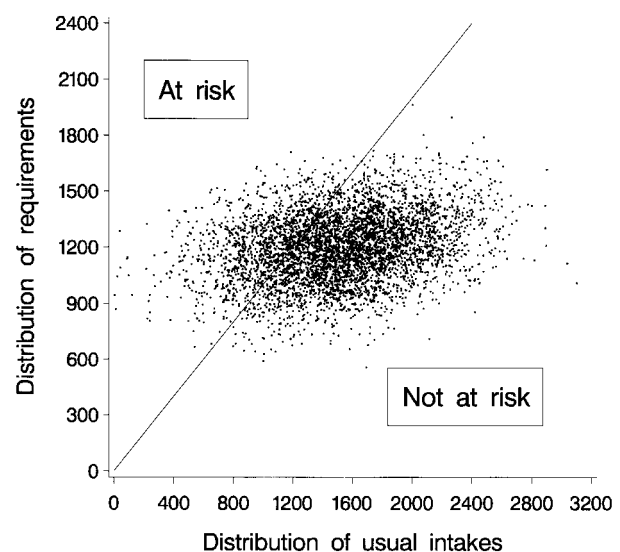


Fig. 1 Simulated joint distribution of intakes and requirements for 5000 individuals. Intakes and requirements are correlated. Along the 45° line, intakes equal requirements

plot presented in Fig. 1. In the figure, we have plotted *simulated* values of requirements and intakes for 5000 individuals. The pairs r_j, i_j were drawn from a (bivariate normal) distribution with mean requirement $\mu_r = 1200$, mean intake $\mu_i = 1500$, 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 modelling nutrient intakes^{1-3,13-16}.

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,

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

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 report¹ or the *cut-point approach* proposed by Beaton⁵ 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 prevalence of nutrient inadequacy becomes tractable. Indeed, as discussed below, methods for reliable estimation of $f_I(i)$ have been proposed², and intake 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^{5,17-21}. 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 the 1986 report¹. 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 Fig. 2 we show the risk curves for two simulated requirements distributions. Both simulated requirements distributions are normal with mean $\mu_R = 42$. The standard deviations of the simulated requirements distributions are 6.5 (solid curve) and 15.0 (dashed 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

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

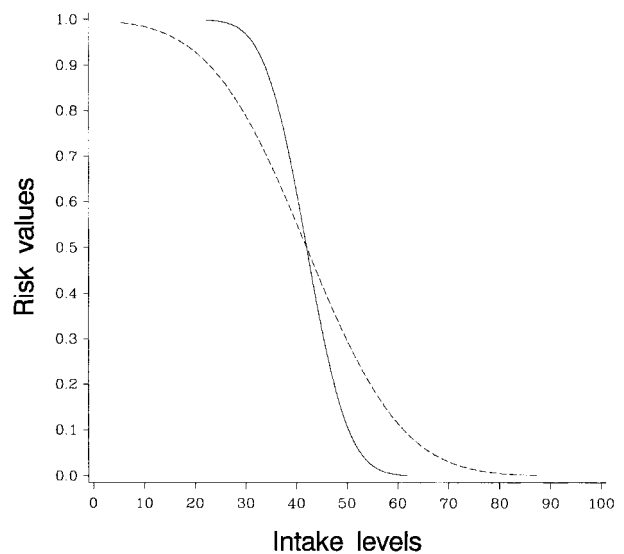


Fig. 2 Two simulated risk curves, obtained assuming that requirements for the nutrient are normally distributed with mean 42 and standard deviation 6.5 (solid line) or 15.0 (dotted line)

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

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 about 100% under both simulated risk curves. An individual with an intake of 50 would have a risk of nutrient inadequacy of about 12% when the standard deviation of the requirements distribution is 6.5 and of about 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 Fig. 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, nobody is at risk of nutrient inadequacy.

We illustrate these two cases in the plots in Figs 3 and 4. In Fig. 3, almost everyone is at some risk, since the distribution of usual intakes in the population says that almost all intakes are between 0 and 80. In Fig. 4, in

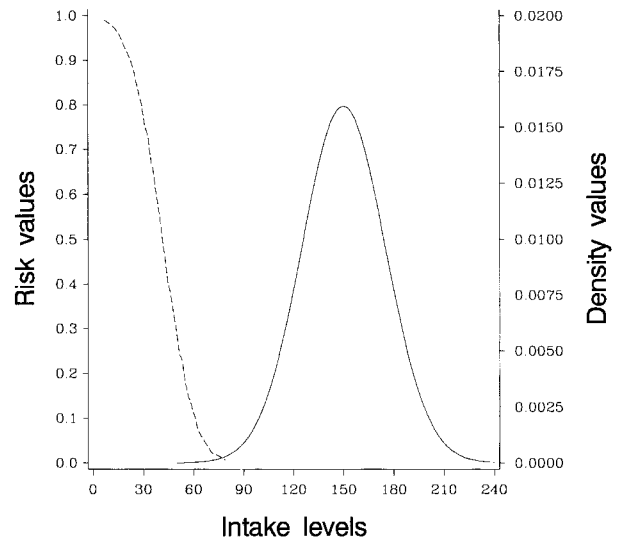


Fig. 4 Simulated risk curve and usual intake distribution. Most of the population is at no risk of nutrient inadequacy

contrast, the prevalence estimate would be about 0%, since 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 Fig. 5. In Fig. 5 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.

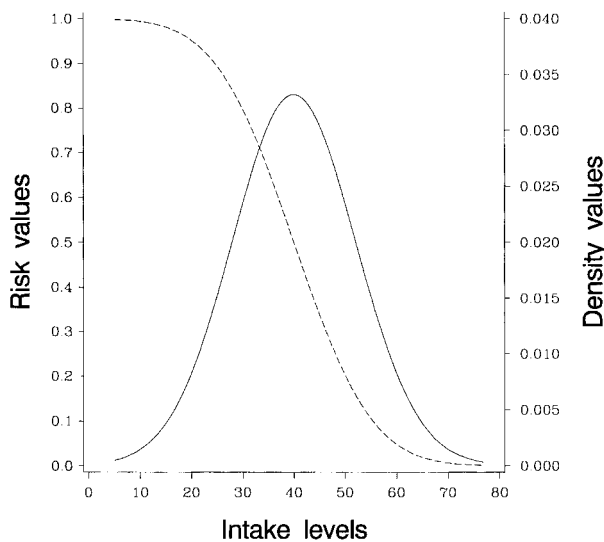


Fig. 3 Simulated risk curve and usual intake distribution. Most of the population is at some risk of nutrient inadequacy

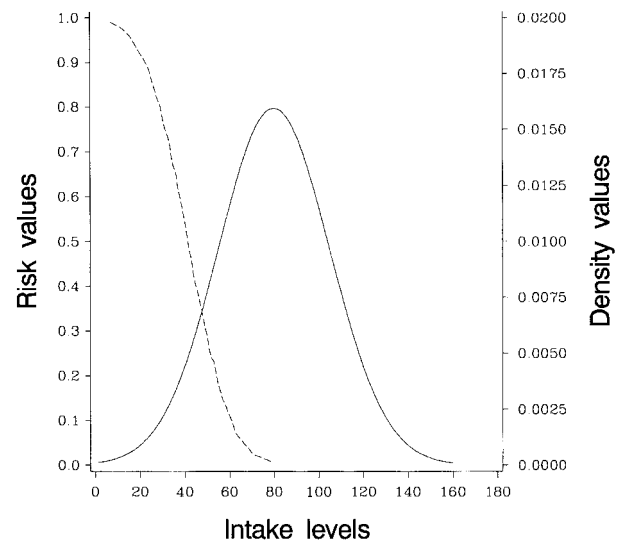


Fig. 5 Simulated risk curve and usual intake distribution. Some individuals in the population are at risk of inadequacy

Notice the following:

- 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 while the mean requirement was set at only 42 for the simulated requirements distribution.
- For an intake level of about 60, the risk of inadequacy is about 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 Fig. 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 up to intake levels with risk of inadequacy of about 0. In the example in Fig. 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 report¹ 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 groups 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 midpoint of the bin. In the example, all intakes between 50 and 60 would have a risk of about 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 ρ associated to the midpoint of each bin and $P(\cdot)$ of the bin. An example of this computation is given on p. 28, Table 5-1, of the NRC report¹.

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 in (2), 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 Figs 2–5 cannot be computed.

In general, researchers assume that requirement 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 report¹ provides various examples of the effect of changing the mean and the variance of the requirement distribution on prevalence estimates. What the NRC report does not mention is that an incorrect specification of the *form* of the requirements distribution may lead to even more elaborate biases.

The EAR cut-point method

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 cut-point method proposed by Beaton⁵ requires no such assumptions, and under certain conditions, should be preferred to the probability approach.

The cut-point method is very simple: estimate prevalence of inadequate intakes as the proportion of the population with usual 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 EAR cut-point method works, we go back to the section where we discussed the joint distribution approach to estimating prevalence. Please refer to Fig. 6, which shows a simulated joint distribution of usual intakes and requirements. To generate the joint distribution, we simulated usual intakes and requirements for 5000 individuals, from independent χ^2 distribution with 7 degrees of freedom and a normal distribution, respectively. The usual intake distribution was re-scaled to have mean 1500 and standard deviation 400. The normal distribution used to represent requirements had mean 1200 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 Fig. 6

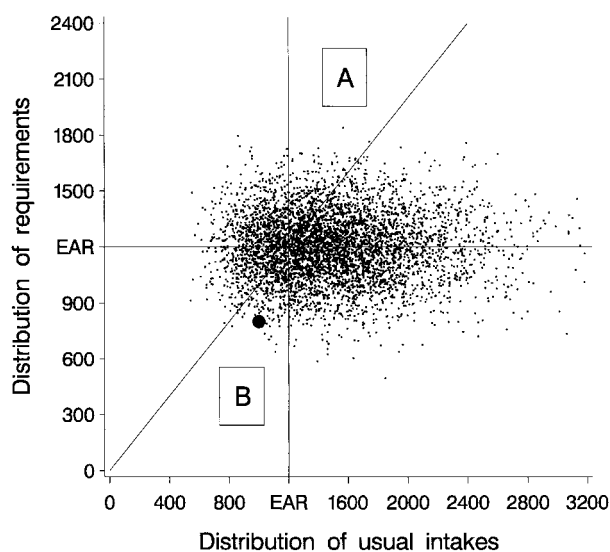


Fig. 6 Simulated joint distribution of intakes and requirements for 5000 individuals. Usual intake distribution is skewed. The black dot represents an individual in the population

has an intake below the mean requirement EAR, but is not at risk since for him/her, $i - r > 0$.

Because we base our inferences on the joint, rather than the univariate, distribution, we may well find that an individual consuming a nutrient at a level below the mean requirement in the population 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 $i = 1200$ line, in the triangle labelled B.

We have agreed that in order to estimate prevalence we would need to ‘count’ those individuals in Fig. 6 who appear *above* the 45° line, since 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 cut-point method is another such approximation, and we now describe its fundamentals.

Refer to the plot in Fig. 6. Notice that the number of individuals with intakes to the *left* of the vertical EAR = 1200 reference line is more or less the same as the number of individuals *over* the 45° line. In other words, 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. The number of individuals in the triangle labelled A appears to be approximately the same as the number in the lower triangle labelled B in Fig. 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 (but not necessarily normal).
- 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 to the assumptions that are implicit in the probability method described above. 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 be known, as is needed 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 the conditions above 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 be large relative to the variance of the requirement distribution. The simulated population shown in Fig. 6 is one for which the three conditions above 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, we would only need reliable estimates of:

- The mean of the requirement distribution for the population, or EAR.
- 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 Fig. 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 above are met (i.e. if our data, were it available, were to 'look' like the data in Fig. 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 assumptions above, $p_{\bar{r}}$ is a good approximation to p_r : those individuals in the top triangle labelled A in the plot, that should be counted but are not, are instead compensated by those individuals in the bottom triangle labelled B in the

plot that should not be counted but are. In other words, under the assumptions above, 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 Fig. 6, the true number of individuals with usual intake below their requirement is 1405. The number that would be estimated using the cut-point approach is 1350. The under-estimation of prevalence in this example is very small.

We conducted a small Monte Carlo experiment to assess the performance of the cut-point approach under varying assumptions for the distributions of usual intakes and of requirements. The results from the simulation experiment are given below.

When the EAR cut-point method does not work well

The cut-point 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 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 of, for example, energy^{1,5,18}, 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 Fig. 7 we show a simulated joint distribution of intakes and requirements obtained by generating 5000 pairs (r , i) from a bivariate normal distribution with mean requirement equal to 1200 and mean intake equal to 1500. 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 bottom triangle labelled B is

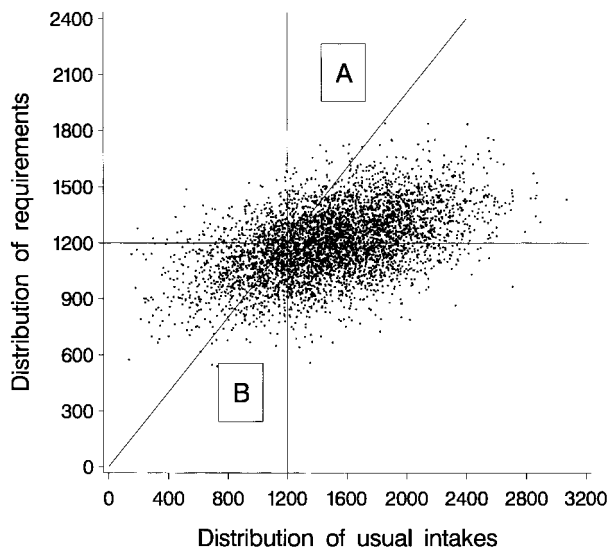


Fig. 7 Simulated joint distribution of intakes and requirements for 5000 individuals. The correlation between r and i is 0.5

larger than the number of points in the upper triangle labelled A. In this case, using the cut-point method would lead to a significant *over-estimation* of the prevalence of inadequate nutrient intakes in the population. In this simulated example, the true prevalence is 21%, while the estimated prevalence using the cut-point method is approximately 27%. This means that in this population of 5000 individuals, we would be over-estimating the number at risk of nutrient inadequacy by about 300 individuals.

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^{1,19}, 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 cut-point approach produces biased prevalence estimates.

Please refer to Fig. 8. In the figure, we show 5000 simulated pairs of intakes and requirements. Here, skewed (χ^2) distributions were used to represent both the intake 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. Thus, 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. In this example, the cut-point approach produces an under-estimation of the true prevalence of inadequate

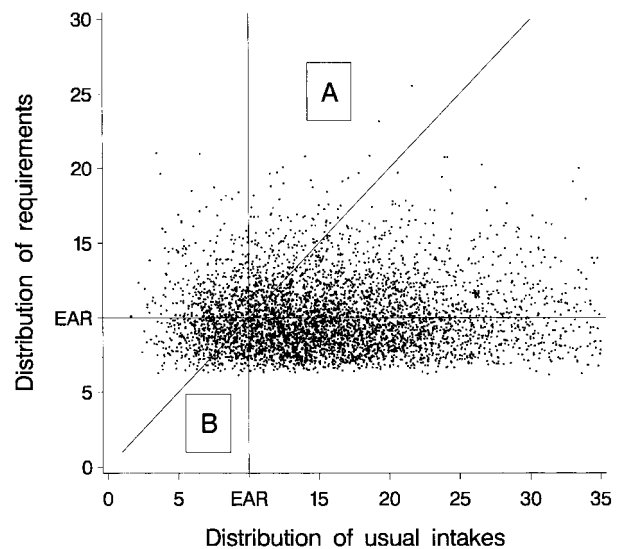


Fig. 8 Simulated joint distribution of intakes and requirements for 5000 individuals. Requirements are assumed to be skewed

intakes. In this simulated example, the extent of the under-estimation is about 200 individuals in the population of 5000.

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 5000 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 EAR cut-point method produces an estimate of prevalence that is noticeably biased relative to what would have been estimated had we had the joint distribution available. In the example shown in Fig. 9, the cut-point method produces a seriously biased prevalence estimate. While the true number of individuals at risk in the population of 5000 is 1414, the cut-point estimate is only 481 individuals.

While the figures we have presented show specific illustrative examples, it is important to note that whether the cut-point method produces an over-estimate or an under-estimate 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, consider the plot shown in Fig. 9. If in this example, we had set the mean intake to be equal to the mean requirement, then the bias of the cut-point approach would have been essentially zero, even though condition 3 above is violated. The effect of

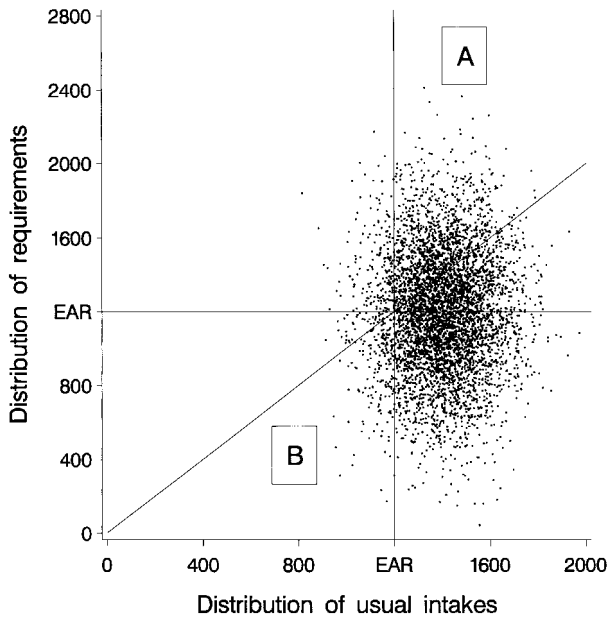


Fig. 9 Simulated joint distribution of intakes and requirements for 5000 individuals. The variance of requirements is large relative to the variance of intakes

the relative sizes of the mean intake and EAR, and of the standard deviation of intakes and requirements are highlighted in the Monte Carlo study presented below.

Evaluating the performance of the EAR cut-point method

How well does the cut-point method perform? We have argued that, as long as the conditions listed above hold, the method works well.

In this section, we present the results of a Monte Carlo study conducted to assess the performance of

Table 1 Expected number of individuals at risk of nutrient inadequacy (in parentheses) and percentage under- or over-estimation of prevalence of short-cut approach

μ_i	σ_r			
	$0.45 \sigma_i$	$0.85 \sigma_i$	$1.45 \sigma_i$	$1.85 \sigma_i$
0.7 EAR	+1.7% (4261)	+5.6% (4066)	+12.4% (3727)	+16.1% (3542)
0.8 EAR	+2.0% (3877)	+6.1% (3671)	+12.1% (3370)	+15.0% (3226)
0.9 EAR	+2.0% (3362)	+5.6% (3184)	+9.8% (2975)	+11.6% (2884)
EAR	+1.5% (2724)	+3.4% (2627)	+4.9% (2552)	+5.4% (2530)
1.10 EAR	+0.1% (2007)	-0.5% (2040)	-2.2% (2124)	-3.2% (2173)
1.20 EAR	-1.8% (1298)	-5.3% (1471)	-10.0% (1707)	-12.3% (1825)
1.30 EAR	-3.4% (705)	-8.8% (976)	-15.8% (1323)	-19.3% (1497)

the cut-point method when condition 3 is not met. We considered 28 different scenarios constructed as follows.

- The EAR was kept constant at a value of 1200.
- The standard deviation σ_i of the simulated usual intake distribution was kept constant at 400.
- The mean of the usual intake distribution μ_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 σ_r was set at four different values: 45%, 85%, 145% and 185% of σ_i .

In all 28 cases the requirements distributions were simulated to be normal with mean EAR and variance σ_r^2 , and the usual intake distributions were simulated as χ^2 distributions with 10 degrees of freedom, re-scaled to have mean μ_i and variance σ_i^2 . Intakes and requirements were simulated to be independent. That is, conditions 1 and 2 above hold in all 28 scenarios.

As in the previous examples, we generated populations of size 5000. For each of the 28 combinations of μ_i , EAR, σ_i and σ_r we simulated 3000 replications of such populations.

Results from the experiment are given in Table 1. In the table we present the *expected* number of individuals at risk of nutritional deficiency in parenthesis. This expectation is computed as the average number, over the 3000 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 4261 (out of 5000).

The expected over- or under-estimation of the proportion at risk that results from using the cut-point 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 3000 replications) of 4333, 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 μ_i and EAR. Similarly, if we move across columns (within a row) we see the effect of the relative sizes of σ_r and σ_i on the performance of the cut-point approach for estimating prevalence.

We note the following:

- The cut-point approach appears to perform well in most situations, as long as $\sigma_r < \sigma_i$. That is, when condition 3 is satisfied (as in the first two columns of the table) then the amount of over- or under-estimation that results from applying the cut-point approach is small.
- When $\sigma_i < \sigma_r$, the biases that result from application of the cut-point approach can be unacceptably high.

For example, the fourth column in Table 1 suggests that we would be under-estimating 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 Fig. 9.

- As expected from statistical theory, the biases approach zero as the mean usual intake approaches the EAR. Please 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 cut-point 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 which suggests that the cut-point approach proposed by Beaton 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 cut-point 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 condition 3 listed above.

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 a nutrient 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^{2,5,22,23}.
- Skewed (not normal or 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^{2,15}.
- 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^{2,3,16}.
- Day-to-day correlation in intakes when dietary data are collected over consecutive days²⁴.
- Complex survey designs that require the use of sampling weights when drawing population

inferences such as estimation of the percent of the population with intakes below (or above) a given cut-point such as the EAR.

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 includes two or more independent 24-h recalls from at least some of the individuals in the sample. If dietary data are collected over consecutive days, then three or more 24-h recalls should be obtained from at least some individuals in the sample. The method was developed by expanding on the procedure proposed in the National Research Council report¹. 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 on that day. Details about the procedure are given in the nutrition³ and the statistics literature^{2,4,16}.

The method developed at ISU² 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 report¹. The performance of the different methods available was tested via simulation experiments². In all cases, the method developed at ISU 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 CSFII from 887 men aged 60 years and older. Estimated distributions are presented in Fig. 10.

The three curves in the plot correspond to estimates of the usual intake distribution obtained by the following three procedures:

- Compute the means of 3 days of intake and fit a distribution to the observed individual means.
- Apply the method proposed in the NRC report¹.
- Apply the method developed at ISU^{2,3}.

Since an EAR for vitamin E is not yet available, we used an arbitrary cut-point of 6 mg α -TE for illustration. It is clear from the plot that the estimate of the proportion of the population with intakes below the (arbitrary) cut-point value of 6 mg α -TE vary greatly depending on the method used to estimate the distribution. Estimated values for the proportion of the population with intake below the cut-point is almost 10% lower in the ISU method than in the NRC procedure. In addition, we estimated the proportion of the population below the cut-point when only the first

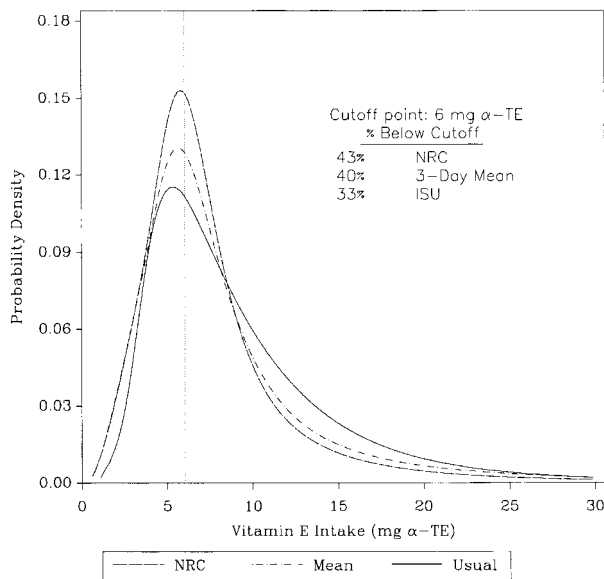


Fig. 10 Estimates of the usual nutrient intake distribution of vitamin E for men aged 60 years and older, using 24-h recall data from the 1989–91 CSFII

day of data is used for each individual in the sample. The estimated proportion below the cut-point 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.

Acknowledgements

This work resulted from discussions among the members of the 'Uses and Interpretations of DRIs' of the Food and Nutrition Board, to whom the author is greatly indebted. 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 Mathematics Department at Iowa State University provided valuable insight about some of the material presented in this paper. This work was partially funded by Contract No. 0009730470 between the National Center for Health Statistics (NCHS) of the Center for Disease Control and Prevention and the Department of Statistics, Iowa State University.

References

- National Research Council. *Nutrient Adequacy: Assessment Using Food Consumption Surveys*. Report of the Food and Nutrition Board. Washington, DC, National Academy of Sciences. 1986.
- Nusser SM, Carriquiry AL, Dodd KW, Fuller WA. A semiparametric transformation approach to estimating usual daily intake distributions. *J. Am. Stat. Assoc.* 1996; **91**: 1440–9.
- Guenther PM, Kott PS, Carriquiry AL. Development of an approach for estimating usual nutrient intake distributions at the population level. *J. Nutr.* 1997; **127**: 1106–12.
- Carriquiry AL, Nusser SM. *Analyzing Dietary Intake Data*. Ames: Iowa State University, Department of Statistics, 1998.
- Beaton GH. Criteria of an adequate diet. In: Shils RE, Olson JA, Shike M, eds. *Modern Nutrition in Health and Disease*. Philadelphia: Lea & Febiger, 1994: 1491–505.
- Murphy SP. The need for different recommendations for individual and population intakes. Unpublished manuscript, University of California, Berkeley, 1994.
- Lörstad MH. Recommended intake and its relation to nutrient deficiency. *FAO Nutr. Newsletter* 1971; **9**: 18–31.
- Peterkin BB, Kerr RL, Hama MY. Nutritional adequacy of diets of low-income households. *J. Nutr. Educ.* 1982; **14**: 102–4.
- Life Sciences Research Office. *Guidelines for use of dietary intake data*. Anderson SA, ed. Bethesda, MD: Federation of American Societies for Experimental Biology, 1986.
- Hegsted DM. Problems in the use and interpretation of the Recommended Dietary Allowances. *Ecol. Food Nutr.* 1972; **1**: 255–65.
- Beaton GH. What do we think we are estimating? In: Beal VA, Laus MJ, eds. *Proceedings of the Symposium on Dietary Data Collection, Analysis, and Significance*. Res. Bull. No. 675, University of Massachusetts at Amherst, 1982.
- Beaton GH. *Empiric relationships between intake and requirement distributions*. Report submitted to the Food and Nutrition Board, Institutes of Medicine, 1995.
- Sempos CT, Johnson NE, Smith EL, Gilligan C. Effects of intraindividual and interindividual variation in repeated dietary records. *Am. J. Epidemiol.* 1985; **121**: 120–130.
- Aickin M, Rittenbaugh C, Woolf K. A general method for adjusting dietary data to obtain stable estimates of usual intake. *Am. J. Clin. Nutr.* 1994; **59**: 303S (abstract).
- Emrich JE, Dennison D, Dennison K. Distributional shape of nutrition data. *J. Am. Diet. Assoc.* 1989; **89**: 665–70.
- Carriquiry AL, Dodd KW, Nusser SM. *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.
- Beaton GH, Chery A. Protein requirements of infants: a reexamination of concepts and approaches. *Am. J. Clin. Nutr.* 1988; **105**: 914–23.
- FAO/WHO/UNU. Energy and Protein Requirements. Report of a Joint FAO/WHO/UNU Expert Consultation. *WHO Tech. Rep. Ser. No. 724*, 1985.
- 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.
- Beaton GH. Recommended dietary intakes: individuals and populations. In: Williams G, Wilkins S, eds. *Modern Nutrition in Health and Disease*, ninth edition. Philadelphia: Lea & Febiger, 1998.
- Dewey KG, Beaton GH, Fjeld C, Lonnerdal B, Reeds P. Protein requirements of infants and children. *Eur. J. Clin. Nutr.* 1996; **50**: S119–S150.
- Borrelli R, Cole TJ, Di Biase G, Contaldo F. Some statistical considerations on dietary assessment methods. *Eur. J. Clin. Nutr.* 1989; **43**: 453–63.
- Beaton GH, Burema J, Rittenbaugh C. Errors in the interpretation of dietary assessments. *Am. J. Clin. Nutr.* 1997; **65**: 1100S–1107S.
- Carriquiry AL, Fuller WA, Goyeneche JJ, Jensen HH. Estimated correlations among days for the combined 1989–91 CSFII. *Dietary Assessment Research Series 4*. CARD Staff Report 95-SR77. Ames: Center for Agricultural and Rural Development, Iowa State University, 1995.

The 8th European Nutrition Conference

17 — 19 June 1999

Radisson SAS Lillehammer Hotel, Lillehammer, Norway

Organisers

Norwegian Nutrition Society (NNS)

Federation of European Nutrition Societies (FENS)

European Association of Nutritional Sciences (EANS)

Organising Committee

Elisabet Helsing, Oslo (conference president)

Sigrid Berge, Oslo (conference secretary)

Kåre Julshamn, Bergen

Arne Løvø, Oslo

Helle M Meltzer, Oslo

Jan I Pedersen, Oslo

Ragnhild Rønneberg, Oslo

Scientific programme

The programme will focus upon nine 'hot' themes from the nutritional sciences:

- Food and nutrition policy
- Nutrition and genes
- Obesity
- Antioxidants and non-nutrients from fruits and vegetables
- Intrauterine and child nutrition
- Nutrition and bone
- Food safety and security
- Micronutrients
- Dietary lipids and health

Important dates and deadlines

Deadline for submission of abstracts 1 March 1999

Registration at discount rate 15 April 1999

For information, contact:

Conference bureau

TS FORUM

P. O. Box 14

N-2601 Lillehammer

Norway

Tel: +47 61 28 73 20

Fax: +47 61 28 73 30

Email: lillar@online.no

The Final Announcement, the Registration Form and the Abstract Application Form are available through the conference homepage at

http://www.nutrition.uio.no/nse/8thFENS_EANS/

