Assessing the Adequacy of Diets: A Brief Commentary

Alicia L. Carriquiry

Briefing Paper 96-BP 11
August 1996

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070
www.card.iastate.edu

This paper is available online on the CARD Web site: *www.card.iastate.edu*. Permission is granted to cite this information with appropriate attribution to the authors.

Questions or comments about the contents of this paper should be directed to Alicia Carriquiry, alicia@iastate.edu.

ASSESSING THE ADEQUACY OF DIETS: A BRIEF COMMENTARY

SUMMARY
In this commentary, we discuss the problem of estimating the proportion of the population at risk of a dietary deficiency. Ideally, we would need an estimate of the proportion of individuals whose intake is below their requirement. This would require knowledge about the *joint* distribution of intakes and requirements in the population. In lieu of information about the joint distribution of intakes and requirements, following Beaton (1993) we argue that it may suffice to know the mean of the requirement distribution together with the distribution of usual intakes. The proportion of individuals with intakes below the mean requirement for the population gives a good approximation to the proportion of individuals with intakes below their requirements, if certain assumptions hold. Those assumptions are: (a) nutrient requirements and usual intakes are independent, (b) the distribution of requirements is symmetric, and (c) the between-individual variance of requirements is small relative to the between-individual variance in intakes.

INTRODUCTION
Assessing the adequacy of diets and estimating the proportion of individuals at risk of dietary deficiency in a (sub)population have been long-standing problems. While different analytical approaches to the problems have been advocated, it appears that, at a minimum, we need information on usual nutrient intake distributions and on nutrient requirements in the population of interest. Today, we have statistical procedures available to obtain reliable estimates of usual intake distributions, and in particular of tail probabilities under those distributions. We present later in this paper a summary of the features of the method developed at Iowa State University to estimate usual nutrient intake distributions (Nusser et al. 1995). Application of these methods for estimating the distribution of usual nutrient intakes requires multiple days of dietary intake

data (two or more if observations are independent) on at least a subsample of the individuals. As has been noted, for example, by Beaton (1993) and Murphy (1994), we do not, in general, have reliable estimates of requirements distributions in the population. As exceptions we have only a very few nutrients in selected sex-age groups.

The information necessary to estimate the joint distributions of requirements and intakes is even more scarce. 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 intake for the jth individual in the population. 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 a dietary deficiency?" 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.

In the remainder of this commentary we briefly discuss the type of data needed to make inferences about the proportion of the population at risk of deficiencies and argue that, under certain assumptions, it may be possible to address this problem with data already available. In the last section, we very briefly present the major attributes of a method for estimating usual intake distributions that was developed at Iowa State University in collaboration with the U. S. Department of Agriculture (USDA). We do not recommend the best procedure for setting

individual or population recommendations. The objectives of this paper are

- To discuss the problem of assessing the proportion of the population at risk of dietary deficiency. In particular, to discuss the method that consists in estimating the proportion of the population with intake below the mean requirement (Beaton 1993).
- To explicitly present the consequences of the assumptions associated with the method proposed by Beaton (1993).
- To discuss the method developed at ISU for estimating usual intake distributions.

COMBINING REQUIREMENTS AND INTAKES FOR RISK ASSESSMENTS

Why would we be interested in estimating the *joint distribution of intakes and requirements*? Consider the plot presented at the top in Figure 1. In the figure, 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 $\mu_i = 1,800$, standard deviation of requirement $\sigma_r = 200$, standard deviation of intake $\sigma_i = 440$, and correlation coefficient between requirement and intake $p_{r,i} = 0.65$. 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 (see Sempos 1985; Aickin and Ritenbaugh 1991).

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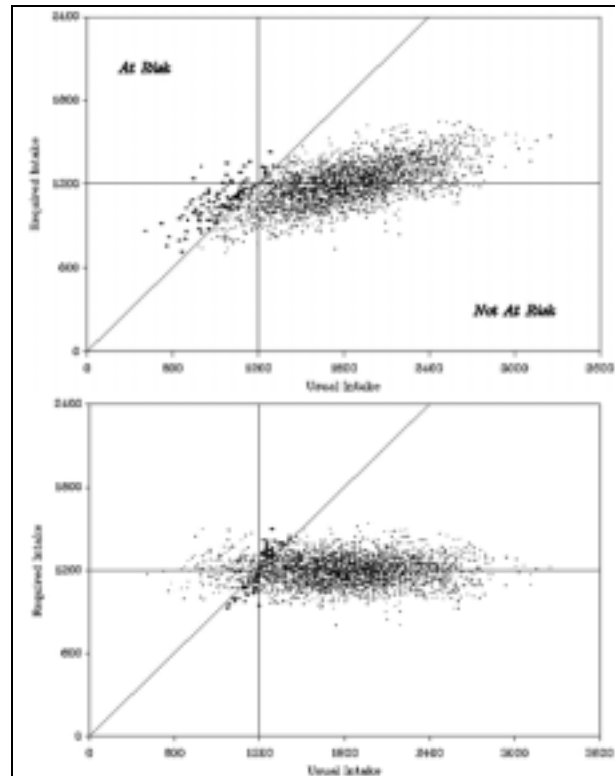

Figure 1: Simulated joint distribution of intakes and requirements for 5,000 individuals. Pairs of observations were drawn from a bivariate normal distribution, with the mean requirement $\mu_r = 1,200$, mean intake $\mu_i = 1,800$, standard deviation of requirement $\sigma_r = 200$, and standard deviation of intake $\sigma_i = 440$.
Top: Correlation between intake and requirement is 0.65.
Bottom: Intakes and requirements are independent.

Notice that under this joint distribution approach, an individual whose intake is below the *mean requirement* is not necessarily at risk. For example, the individual marked with a "+" in Figure 1 has an intake below the mean requirement but is not at risk since for him/her $i - r > 0$. 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 not be at risk. That is the case with all the individuals in the plot who appear below the 45° degree line and to the left of the $i = 1,200$ line.

Unfortunately, we do not have, at present, the information needed to obtain reliable estimates of joint distributions of requirements and intakes. In this light, Beaton (1993) and Murphy (1994) have

advocated different methods to estimate the proportion of the population at risk and thus set both individual and population recommendations. Here, we consider the method proposed by Beaton (1993), which consists of estimating the proportion of the population at risk by calculating the proportion of individuals with intakes below the mean population requirement.

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:

- Intakes and requirements are independent. In statistical terms, this implies that their joint distribution can be estimated as the product of the marginal distributions.
- The distribution of requirements is symmetric.
- 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.

In the plot at the bottom of Figure 1 we show 5,000 simulated pairs of requirements and intakes, generated from a bivariate normal distribution with the same means and variances as before, but where the correlation between intake and requirement is equal to zero. In the normal case, zero correlation implies independence of intake and requirement. Notice that these data have been generated so that the assumptions listed above are all 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, as was pointed out by Beaton (1993), we would only need reliable estimates of

- The mean of the requirement distribution for the population.
- 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 at the bottom in Figure 1 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 Figure 1), 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_{mr}. Note that under the assumptions above, p_{mr} is a good approximation to p_r: those individuals in the top shaded triangle in the plot, who should be counted but are not, are instead compensated by those shaded individuals in the bottom triangle 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° degree line can be approximated by the proportion to the left of the $i = mr$ (intake = average requirement) line.

What would happen if the assumption of independence of intakes and requirements did not hold? 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. Consider the simulated set of individuals shown in the top plot of

Figure 1. We cannot use the proportion of the population to the left of the $i = 1,200$ line to approximate the proportion above the 45° degree line; when requirements and intakes are not independent, the proportion of individuals in each of the two triangles defined by the $i = 1,200$, the 45° degree, and the $r = 1,200$ can be very different. In our example, there are many more individuals with intakes below the mean requirement but above their own requirement than with intake above the mean requirement but below their own requirement. Thus the approximation is an overestimate of the proportion at risk when the correlation is positive.

It is important to notice the following:

- We cannot make individual-level recommendations like the traditional Recommended Dietary Allowances based on this approximation. To make individual recommendations, we require knowledge of the entire requirements distribution, not just of its mean. (Knowledge about the mean and standard deviation of the requirement distribution would suffice under the assumption of normality.)
- While assessing the proportion of the population at risk appears to be an approachable problem, it is not clear how this information would be used to make recommendations for the population. For example, population recommendations might be based on shifting the mean of the usual intake distribution so that only 2 to 3 percent of the population has intakes below the mean requirement. Implicit in this procedure is the assumption that the *variance* of the intake distribution does not change when we shift its mean. This assumption is likely to be false. Thus, a more realistic approach to making population recommendations would either account for changes in both mean and variance of the intake distribution or else consider shifts in the mean intake that would keep constant the coefficient of variation of the distribution.

ESTIMATING USUAL INTAKE DISTRIBUTIONS

It is apparent from the discussion in the previous section that reliable estimates of usual intake distributions are needed to assess the proportion of the population at risk of dietary deficiency.

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.
- Skewed (not normal nor symmetric) intake distributions. In some nutrients, there appears to be a very long tail to the right of the distribution.
- 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.
- Day-to-day correlation in intakes when dietary data are collected over consecutive days.
- 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 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 includes 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 National Research Council (NRC 1986) report. 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.

The method developed at ISU 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 (1986) report. The performance of the different methods available was tested via extensive simulation experiments, and results can be found in Nusser et al. (1995). 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 Figure 2.

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

- Computing the means of three days and fit a distribution to the observed individual means.
- Applying the method proposed in the NRC (1986) report.
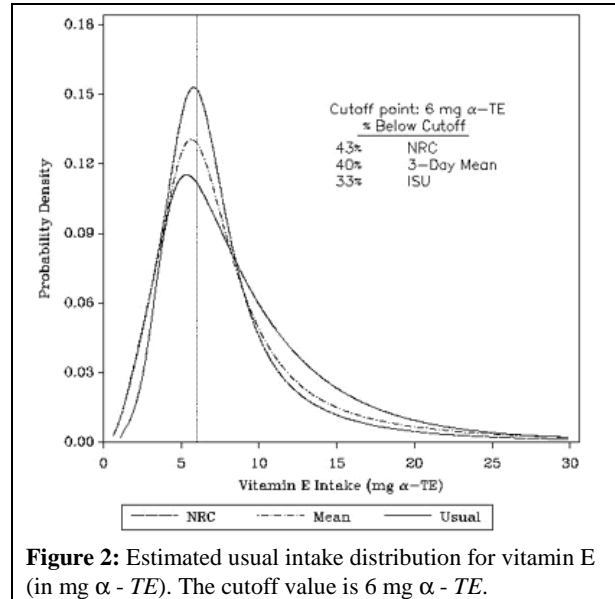- Applying the method developed at ISU (Nusser et al. 1995).

Figure 2: Estimated usual intake distribution for vitamin E (in mg α - *TE*). The cutoff value is 6 mg α - *TE*.

It is clear from the plot that the estimate of the proportion of the population with intakes below the (arbitrary) cutoff 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 cutoff is almost 10 percent lower with the ISU method than with 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 percent, 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

Aickin, M., and Ritenbaugh, C. 1991. Estimation of the True Distribution of Vitamin A Intake by the Unmixing Algorithm. *Communications in Statistics-Simulations* 20: 255-280.

Beaton, G.H. 1993. Criteria of an Adequate Diet. In: Shils, M.E., Olson, J.A., and Shike, M. (eds.) *Modern Nutrition in Health and Disease*, 8th edition. Philadelphia: Lea & Febinger.

Murphy, S.P. 1994. The Need for Different Recommendations for Individual and Population Nutrient Intakes. Unpublished manuscript. University of California, Berkeley.

National Research Council. 1986. *Nutrient Adequacy*. Washington, D.C.: National Academy Press.

Nusser, S.M., Carriquiry, A.L., Dodd, K.W., and Fuller, W.A. 1995. A Semiparametric Transformation Approach to Estimating Usual Intake Distributions. Tentatively accepted by the *Journal of the American Statistical Association*. Also available as Preprint #95-41, Department of Statistics, Iowa State University, Ames, IA 50011-1210.

Sempos, C.T., Johnson, N.E., Smith, E.L., and Gilligan, C. 1985. Effects of intraindividual and interindividual variation in repeated dietary records. *American Journal of Epidemiology* 121: 120-130.