

Background Information
IOM/FNB Workshop on Dietary Reference Intakes
The Development of DRIs 1994-2004: Lessons Learned & New Challenges
September 18-20, 2007
Washington, DC

Information Compiled and Posted July 11, 2007

Purpose: To Provide Useful/Relevant Information for
Workshop Participants and Attendees

Opportunity for interested parties to comment electronically through August 11, 2007:
www.iom.edu/driworkshop2007

DOCUMENT:

Issues of Variability

Developed by:
William Rand, PhD
Professor Emeritus
Department of Public Health and Family Medicine
Tufts University School of Medicine

Date: June 2007

VARIABILITY

The DRI process is directed toward answering the question: “How much should people eat?” The process examines nutrients one by one, and for each identifies associated adverse health effect(s) (AHE), evaluates the relevant data, and synthesizing these data into recommendations. One of the most difficult problems encountered this task is the fact that the data are variable – measured human response to food is not consistent -- different individuals respond to the same intake levels differently, the same individual responds differently at different times, and replicate samples often do not closely agree.

In general, variability is handled by the tools of probability and statistics. This starts with the recognition that the minimum intake level an individual needs (requirement), as well as the maximum that an individual can be tolerated (toxic level), are random variables. Random variables are described by probability distributions (risk curves) that summarize the information relating specific levels of intake with specific risks (chance or probability).

The risk curve describes biological response and is primarily an object of scientific interest. Combined with intake data (real or hypothetical) it provides a wide range of information about the actual or expected prevalence of health problems in either an individual or groups. These prevalence curves can be used for the full range of activities from assessing the diets of individuals to the planning of diets to be offered to large groups.

VARIABILITY

We cannot guarantee specific nutrient levels to be safe for every one. What is sufficient for one individual is inadequate for another, what is sufficient at one time is not enough at another. Similarly one person can tolerate higher levels than can another person or even the same person at another time. This variability of response is a complex interaction between genotype and phenotype, biology and environment, present and past, and moreover is inflated by the errors involved with measurement. While it is possible to adjust somewhat for basic biological factors such as gender, age, and body size, data on similar individuals still differ and replicate measurements on the same individual at different times are not the same, and duplicate samples on the same individuals at the same time themselves often do not agree.

An individual’s requirements and toxic levels are random variables – they cannot be known exactly and the best that can be done is to say where they are likely to be. A major task of the DRI process is to summarize the existing data in a way that estimates the likelihood (risk) of an AHE associated with any specific level of intake of a specific nutrient.

Information about any random variable is summarized by its distribution, a curve that relates each level of the random variable with the probability of that level. **Figure 1** shows the probability distribution of protein requirement, it is centered at 0.66 (the EAR), most individuals lie below 0.8 (the RDA) and the distribution is skewed. For any level of intake the area under the curve to the right of that intake is the probability (risk) of an individual who consumes this level is consuming below requirement. Equivalently, this area estimates the risk (percentage) of a population, fed at that level, would consume less than requirement. While this curve has a complex formula, modern computers can easily calculate it and display it graphically (**Appendix 1**).

THE RISK CURVE

The risk or dose response curve is the cumulative probability distribution. The cumulative distribution is more immediately useful than the probability distribution itself in that risks can be read or calculated directly from the curve. For each intake level the curve provides an estimate of the probability (risk) of that level not meeting an individual's requirement. Equivalently, for each intake level, the curve shows just what fraction of the population has requirements that are below that level. The curve can also be used in reverse -- it can be entered with a risk level and the corresponding intake level found (see **Figure 2** for the risk curve for protein).

The risk curve summarizes data relating levels of nutrient intake and health effects. Like the probability distribution it graphically shows the essential characteristics of a random variable, its center (the mean or median of requirement), its spread (the amount of variability of the response, essentially the slope of the curve near the median), and its shape (basically whether skewed or not). Usually the risk curve for the requirement of a nutrient is derived from its probability distribution and based on estimates of mean/median, the standard deviation/coefficient of variation, and the shape. The risk curve exists both as an easily read graph and as a, usually complex, mathematical formula. The necessity of working with random variables in the real world has produced easy computer tools to work with most of these formulae that are readily available. (**See Appendix 1.**)

The risk curve stands at the center of the Dietary Reference Intakes process. Much of the effort of the DRI panels is directed to finding and using the available data on intake and health effects to estimate a risk curve for each nutrient and adverse health effect. However, if insufficient data are available no risk curve can be estimated, and alternative procedures are necessary (calculation of the AI, etc.)

ESTIMATING THE RISK CURVE

The risk curve is estimated from experimental data and in general, the more data available the better estimate. The important characteristics of the risk curve are its shape, its center, and its spread. These can all be determined if there are enough data, however

usually certain assumptions need be made. In general one needs only a little data to guess at the center of a distribution, one needs more data to estimate the spread of the distribution, and a lot of data to be confident of the shape of the distribution.

Table: General Risk Curve Estimation

DATA AVAILABLE	RISK CURVE ESTIMATION PROCEDURE
Lots of data over full range of intake	Estimate shape, Estimate mean/median (EAR), Estimate standard deviation (SD) or coefficient of variation (CV)
Some data over the full range	Estimate mean/median, Estimate SD or CV Assume shape
Little data, all in the middle of range	Estimate mean/median Assume SD or CV Assume shape
Insufficient data	Use alternative risk summary (e.g., AI)

It has been found that many important biological distributions can be satisfactorily represented as either Normal (if symmetric) or Log-Normal (if skewed). Because of this, the DRI process assumes a Normal distribution when the requirement is not noticeably skewed and usually Log-Normal distribution when skewing is present (although for Iron the distribution was simulated). If there is not good information of the spread, a standard deviation or coefficient of variation is estimated based on other, related data. If no center can be estimated, the process is abandoned and other sorts of data explored, most often the intake of a healthy population (giving an AI). The following examples show these procedures:

EXAMPLE 1: Adult Protein Requirement.

Protein requirement data exist for 225 “representative” adults so that the risk curve could be estimated directly from the frequency distribution. Statistical examination showed that requirement was not related to either age or gender (for adults) in that they did not reduce the variability consistently or significantly and therefore all the data could be pooled. Examination of the histogram of requirement, adjusted for weight, showed that it could be satisfactorily represented by a log-normal distribution (in the sense that the log of protein requirement was distributed normally). The median and spread were estimated from the data and used to estimate the probability distribution that was then integrated to give the risk curve. (See Appendix 1 and Figure 1 & 2).

Example 2: Adult Vitamin A requirement.

Here, sufficient data were not available to estimate the shape of the distribution, however, data could be used to locate the center (mean 625 for men and 500 for women) and related data suggested a coefficient of variation of 20%. Normality was assumed and the two risk curves shown in figure 3 were estimated. (See Figure 3.)

Example 3: Adult Zinc requirement.

For Zinc, the center could be estimated (9.4 for men and 6.8 for women) but too little data exist to estimate the spread and a coefficient of variation of 10% assumed. Since

there was no information about the shape of the distribution Normality was assumed. (Figure 4.)

Example 4: Adequate Intake

For a number of nutrients (e.g., Vitamin K and Chromium) neither the center nor spread could be estimated and so the intake of a healthy population was used to estimate intake assumed to be adequate.

The risk curve summarizes the scientific data relating intake level to AHE. It can be represented both graphically and mathematically and can be used both to estimate the risk that is associated with a specific intake level and to estimate the intake level that is associated with a particular risk. By itself the risk curve is of primarily scientific interest; its importance is that it can be used to estimate the prevalence (incidence) of Adverse Health Events in real world situations, where variability in intake is observed or expected.

USE OF THE RISK CURVE

The primary use of the risk curve is to estimate, for an individual or a population, the chance (prevalence, or incidence) of an AHE due to a specific nutrient. This requires information on the nutrient intake of the individual or group in addition to the risk curve for that nutrient. Since both intake (of individual or group) and requirement are (usually) random variable, these distributions can be combined to calculate the desired prevalence. Moreover, in addition to estimating prevalence for known intake distributions, the effect of hypothetical intake distributions can be explored to find intakes levels which will produce specific levels of prevalence.

For a constant diet the risk of an AHE can be calculated directly from the risk curve as the risk value associated with the specified intake level. In the same way, to get a specific intake level related to a specific level of risk one reads the curve in the other direction, entering the acceptable risk level and reading the intake expected to achieve it. This gives estimates of the effects of diets with fixed intake levels.

However, at least in the developed world, very few individuals or populations consume fixed level diets. Dietary intake is a random variable, with a distribution that can be estimated in much the same way that requirement distribution is estimated – from samples, related data, and reasonable assumptions. The distribution of intake can then be combined with the requirement distribution to estimate AHE prevalence.

The prevalence for an AHE is the chance that an individual consumes less than that individual requires -- that individual intake is below individual requirement. Since the difference between two random variables is itself a random variable, this difference (nutrient deficit) has a distribution. This risk of an AHE (its prevalence) is the probability that this difference is negative. Graphically the area to the right of zero on the

distribution, or the prevalence related to zero deficit on the cumulative distribution.
(See Appendix 2)

In general the distribution of the difference between two random variables is a complicated mathematical expression, however, as in the case of the risk curve, there are computer functions and programs available to do the actual calculations. Moreover, since most requirement data follow a normal or lognormal distribution, as do many intakes, the difference distribution can be easily estimated and used to get the probability of the prevalence/incidence. (See Appendix 2.) (Note that procedure is a formalization and extension of the cut-point approximation.)

SUMMARY

Variability is a major feature of human responses to dietary intake, and a major problem in formulating recommendation of dietary reference intakes. Dietary requirements and toxicities are random variables, and information about them is summarized by their distributions and the equivalent risk curves. For each nutrient and AHE the risk curve can be combined with a distribution of real or theoretical intakes to provide estimates of prevalence (incidence) of the AHE that potential users can use to evaluate and plan diets and food programs.

The estimation of risk curves is often not straightforward, hampered by a number of realities, most obvious of which is lack of good data. Without lots of data, assumptions must be made about distributional shape, without data on the variability of the response assumptions must be made for that, and without good estimates of the center of the distribution the estimation process must be given up entirely and alternatives recommendations made, such as the AI.

The complexity of the calculations involved is no longer a problem in the estimation of risk curves. Explicit consideration of both requirement and intake as random variables permits the use of statistical techniques that have been developed for other fields. These call for a certain amount of manipulations of often of fairly complex formulae. However, modern computers both make these calculations quite easy and provide useful graphical displays.

APPENDIX 1: Calculating the probability and cumulative probability distribution

The **probability distribution** of a random variable that is Normally distributed with mean m and standard deviation s , is described by the following equation:

Normal distribution:
$$\frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{(x-m)^2}{s^2}\right]$$

As forbidding as this looks, it is easy to calculate the this distribution for any level of intake “ x ”, using, for example, the built-in function “NORMDIST” in the Microsoft Excel spreadsheet program.

For example, the Zinc requirement for adult males is normally distributed with mean 9.4 and standard deviation 0.94 (a coefficient of variation of 10%). The curve can be calculated for any intake level by:

$$\text{NORMDIST}(\text{Intake Level}, 9.4, 0.94, 0).$$

The distribution curve can be plotted by iterative use of this function for a sequence of intake points over the range of interest.

The **cumulative distribution** of a random variable that is Normally distributed with mean m and standard deviation s , is described by the following equation (for an arbitrary intake level A):

Cumulative Normal Distribution:
$$\int \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{(x-m)^2}{s^2}\right] dx$$

This is the risk curve, and the risk of any level of intake is also easily evaluated using the Microsoft Excel NORMDIST function.

For example, the adult female requirement for Zinc has a normal distribution with a mean of 6.8 and a standard deviation of 0.68. The risk associated with any Zinc intake level is:

$$1 - \text{NORMDIST}(\text{Intake Level}, 6.8, 0.68, 1).$$

The risk curve can be plotted by iterative use of this function for a sequence of intake points over the range of interest.

Notes:

1) In general, the distribution and risk curve for a skewed variable, that can be approximated by a Log-Normal distribution, is calculated by using the above formulae with the log of the variable of interest.

2) Additionally, if enough information is available to be sure that requirement is neither normal nor log-normal, these probability distributions can often be simulated.

APPENDIX 2: Estimating prevalence

Prevalence (of an AHE) is defined as the probability that intake is less than requirement. If two random variables are Normally distributed then their difference is also Normally distributed. The difference has a mean (M) that is the difference between the two component means and a standard deviation (SD) that is the square root of the sum of the two standard deviations squared.

$$\begin{aligned}\text{Mean (intake - requirement)} &= M(\text{intake}) - M(\text{requirement}) \\ \text{Standard Deviation (intake - requirement)} \\ &= \text{Square Root } [SD^2(\text{intake}) + SD^2(\text{requirement})].\end{aligned}$$

From this basic rule of probability it is possible to calculate both the probability distribution and cumulative distribution of the difference between intake and requirement and therefore the probability that is negative. The equations follow those of Appendix 1, and can be easily evaluated using computer functions as shown below.

Individual Example – Chance of intake being below requirement

Adult female Vitamin A requirement is normal with mean of 500 and standard deviation of 100 (20% CV). (See Figure 5a) The distribution is generated by:

$$\text{Requirement: } \text{NORMDIST}(A, 500, 100, 0).$$

A woman with an average intake of 700 and daily variability (SD) of 200 has an intake distribution (See Figure 5a) that is generated by:

$$\text{Intake: } \text{NORMDIST}(A, 700, 200, 0).$$

The distribution of her nutrient deficit is generated by (see Figure 5b):

$$\text{Deficit: } \text{NORMDIST}(A, 200, 224, 0) = 18.6.$$

(where the mean 200 is 700-500, and the SD 224 is the square root of $(100^2 + 200^2)$).

The chance that she is consuming less than her requirement the area under this curve to the left of zero, and is calculated from the cumulative distribution as:

$$\text{NORMDIST}(0, 200, 224, 1) = 0.186.$$

This says that a woman with this intake is has almost a 20% chance of being Vitamin A deficient.

Group Example - Expected prevalence of Vitamin A deficiency of a group of adult men with a specific diet

Adult male Vitamin A requirement is normal with mean of 625 and standard deviation of 137.5 (20% CV). The distribution is generated by:

$$\text{Requirement: } \text{NORMDIST}(A, 625, 137.5, 0).$$

A hypothetical diet which provides a mean intake of 1000 with a daily variability (SD) of 500 has an intake distribution generated by:

$$\text{Intake: } \text{NORMDIST}(A, 1000, 500, 0).$$

The distribution of Vitamin A deficiency in a group consuming this diet is generated by:

$$\text{Deficit: } \text{NORMDIST}(A, 375, 519, 0),$$

(where the mean 375 is 1000-625, and the SD 519 is the square root of $(137.5^2 + 500^2)$).

The chance that she is consuming less than her requirement the area under this curve to the left of zero, and is calculated from the cumulative distribution as:

$$\text{NORMDIST}(0,375,519,1) = 0.235.$$

This says that supplying a group with this diet is likely to result in an almost 25% prevalence of Vitamin A deficiency.

Notes:

1) The above is based on the assumption of a Normal distribution, as has been assumed for many nutrient requirements and intakes. If a log-normal distribution is assumed (eg Protein) the same calculation apply, using the logs of requirement and intake.

2) The above assumes that intake and requirement are not correlated. If a correlation exists, and can be estimated then the deficit standard deviation must be corrected:

$$\begin{aligned} \text{SD (intake - requirement)} &= \text{Square Root} \left[\text{SD}^2(\text{intake}) \right. \\ &\quad \left. + \text{SD}^2(\text{requirement}) \right. \\ &\quad \left. - 2 * \text{correlation} * \text{SD}(\text{intake}) * \text{SD}(\text{requirement}) \right]. \end{aligned}$$

3) The above is an extension and formalization of the “cut point” method.

REFERENCES:

Much of the material presented in this paper is drawn from two major sources; expositions of basic statistics/probability and previous WHO/FAO/(UNU) and NAS reports.

Basic Statistics and Probability

Mood AM, Graybill FA, and Boes DC, (1974) Introduction to the Theory of Statistics, 4th Ed.. Mc Graw Hill.

Hogg RV, Craig A, McKean JW. (2004) Introduction to Mathematical Statistics, 6th Ed. Prentice Hall.

Johnson NL, Kotz S, Balakrishnan N. (1994) Continuous Univariate Distributions, Volume 1, 2nd Ed. John Wiley & Sons, New York, NY.

Kotz S, Balakrishnan N and Johnson NL. (2000) Continuous Multivariate Distributions, Volume 1: Models and Applications, 2nd Ed.

Ramsey FL and Schafer DW. The Statistical Sleuth, 2nd Edition. Duxbury, Pacific Grove, California. 2002.

Afifi AA, Clark V, and May S. (2004) Computer-Aided Multivariate Analysis. 4th Edition. Chapman & Hall, New York, NY

“Official” Reports:

FAO/WHO/UNU. (1985) Energy And Protein Requirements. Report of a Joint Expert Consultation. World Health Organ Tech Rep Ser. Geneva.

National Research Council. (1989) Recommended Dietary Allowances, 10th Ed. National Academy Press, Washington DC.

Institute of Medicine. (1998) Dietary Reference Intakes: a risk assessment model for establishing upper intake levels for nutrients. National Academy Press, Washington, DC.

Institute of Medicine. (2000) Dietary Reference Intakes: Applications In Dietary Assessment. National Academy Press, Washington, DC.

Institute of Medicine. (2001) Dietary Reference Intakes: Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington, DC.

Institute of Medicine. (2003) Dietary Reference Intakes: Applications In Dietary Planning. National Academy Press, Washington, DC.

Institute of Medicine. (2003) Dietary Reference Intakes: Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. National Academy Press, Washington, DC.

Additional details and alternative presentations

Buenther PM, Kott PS, and Carriquiry AL. (1997) Development of an approach for estimating usual nutrient intake distributions at the population level. J. Nut 127: 1106-1112.

Carriquiry AL (1999) Assessing the prevalence of nutrient inadequacy. Publ Health Nutr 2:23-33.

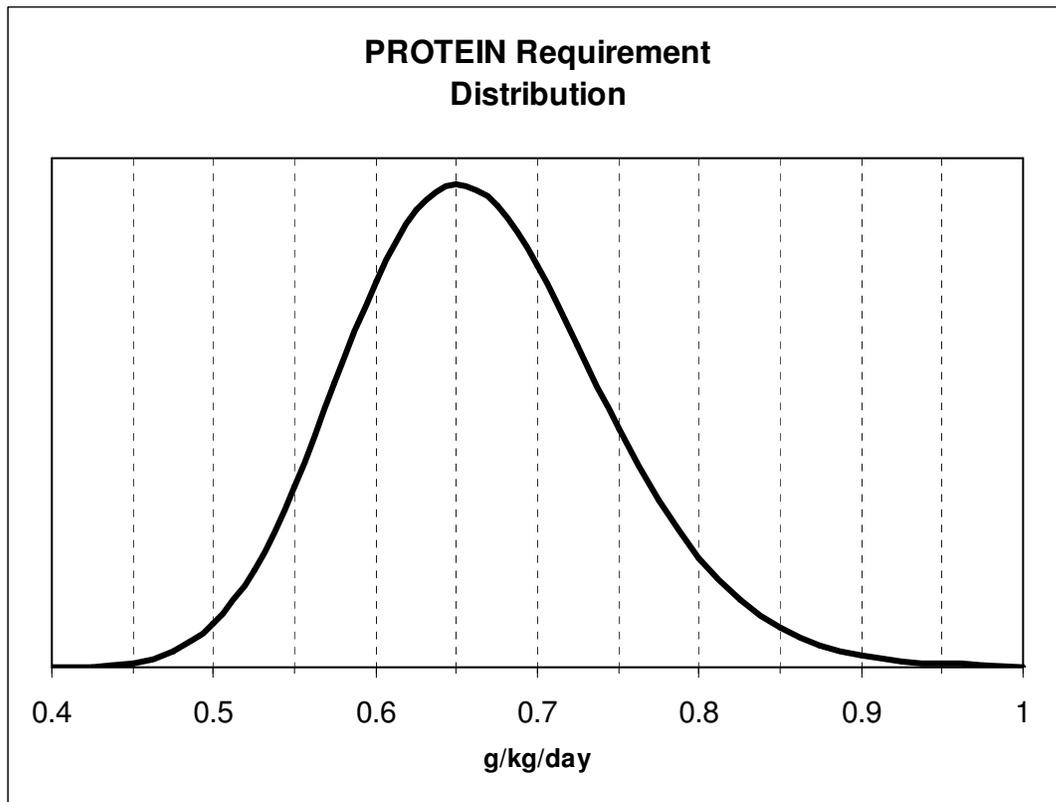
Carriquiry AL. (2003) Estimation of usual intake distributions of nutrients and foods. *J. Nutr.* 133: 601S-608S.

Nusser SM, Carriquiry AL, Dodd KW and Fuller WA. (1996) A semiparametric transformation approach to estimating usual daily intake distributions. *J Am Stat Assoc.* 91(436): 1440-1449.

Rand WM. (2004) Variability in planning and assessment: Statistical considerations and models. *J. Nut* 134: 1610S-1616S.

Rand WM and Young V R (2003) Meta-analysis of nitrogen balance studies for estimating protein requirements in healthy adults. *Am. J. Clin. Nutr.* 77:109-127.

FIGURE 1



Probability Distribution of Adult Protein Requirement.

Curve is the exponential function of a normal distribution with Mean -0.425 and Standard Deviation 0.12. Normality, mean, and standard deviation all based on analysis of 225 individual data points.

Curve was generated using Microsoft Excel formula:

$$\text{NORMDIST}(A, -0.43, 0.12, 0)$$

where A contains values for the natural log of values of intake levels.

(For details see: Institute of Medicine. (2003) Dietary Reference Intakes: Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. National Academy Press, Washington, DC.)

FIGURE 2

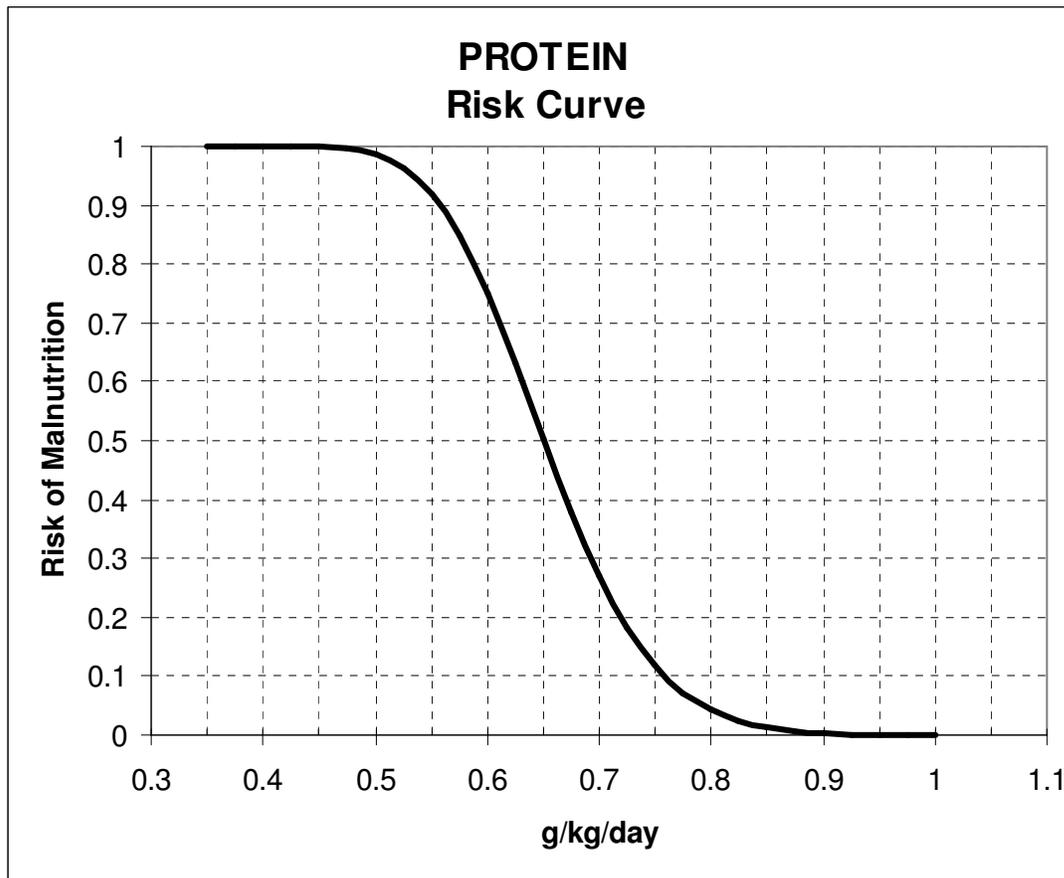


Figure 2: Cumulative Distribution (Risk Curve) of Adult Protein Requirement.

Curve is the cumulative exponential of a normal distribution with Mean -0.425 and Standard Deviation 0.12 , where these values were based on analysis of 225 individual data points

Curve was generated using Microsoft Excel formulae:

$$1.0 - \text{NORMDIST}(A, -0.43, 0.12, 1)$$

where A contains values for the natural log of values for intake

(For details see: Institute of Medicine. (2003) Dietary Reference Intakes: Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. National Academy Press, Washington, DC.)

FIGURE 3

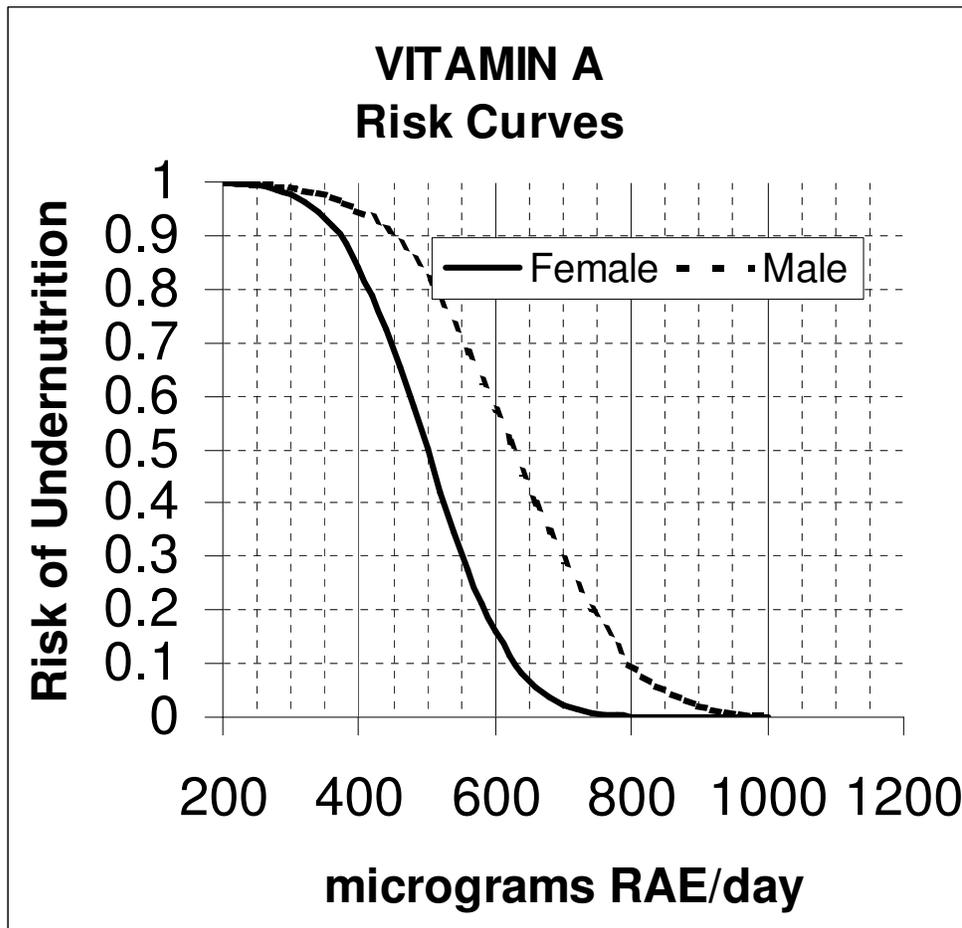


Figure 3: Cumulative Distribution of Adult Vitamin A Requirement.

Curve for females is a normal distribution with Mean 500 and Standard Deviation 125.

Curve for males is a normal distribution with Mean 625 and Standard Deviation 150.

Means and standard deviations from data, Normality assumed.

Curves were generated using Microsoft Excel formulae:

$$\text{Risk for females} = 1 - \text{NORMDIST}(A, 500, 100, 1)$$

$$\text{Risk for males} = 1 - \text{NORMDIST}(A, 625, 137.5, 1)$$

where A contains values for intake.

(For details see: Institute of Medicine. (2001) Dietary Reference Intakes: Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington, DC.)

FIGURE 4: ZINC

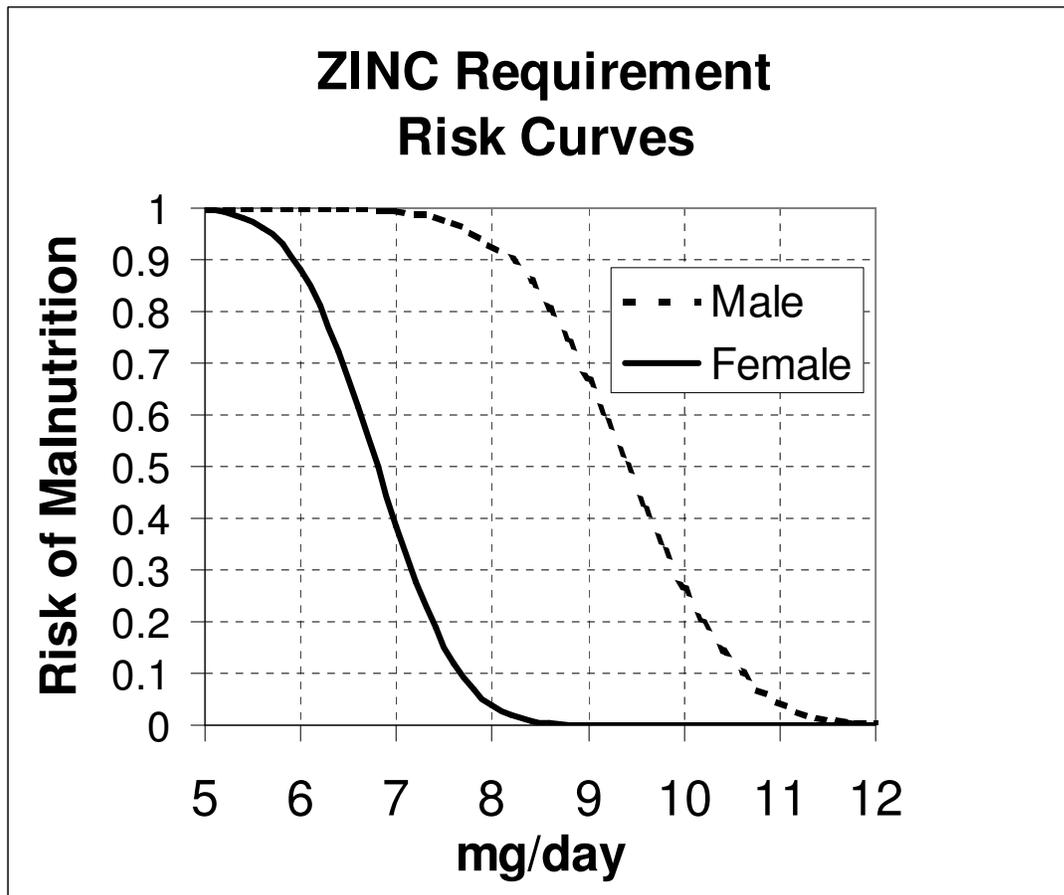


Figure 4: Cumulative Distribution of Adult Zinc Requirement.

Curve for females is a normal distribution with Mean 6.8 and Standard Deviation 0.68.

Curve for males is a normal distribution with Mean 9.4 and Standard Deviation .94.

Means were from data, standard deviations from assumed (10% coefficient of variation), Normality was assumed

Curves were generated using Microsoft Excel formulae:

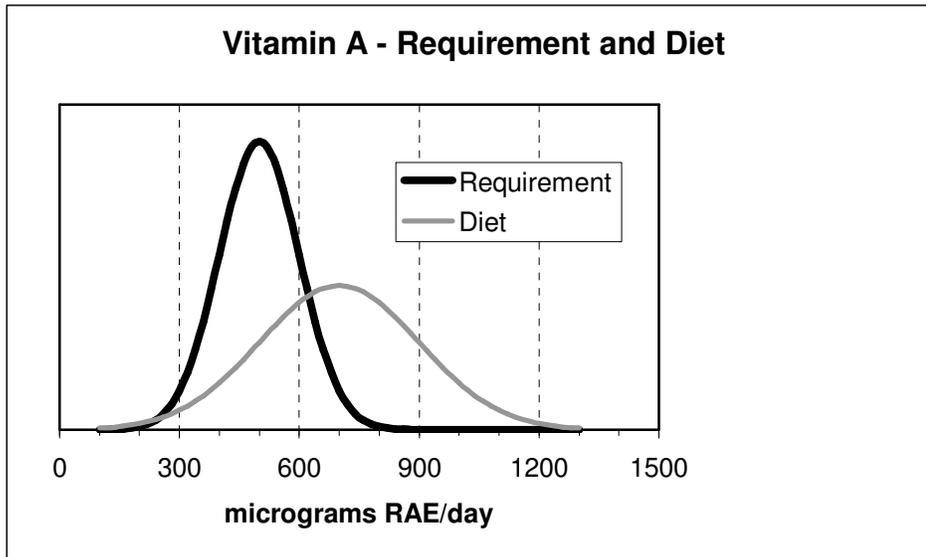
$$\text{Risk for females} = 1 - \text{NORMDIST}(A, 6.8, 0.68, 1)$$

$$\text{Risk for males} = 1 - \text{NORMDIST}(A, 9.4, 0.94, 1)$$

where A contains values for intake

(For details see: Institute of Medicine. (2001) Dietary Reference Intakes: Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. National Academy Press, Washington, DC.)

FIGURE 5a: Vitamin A Example



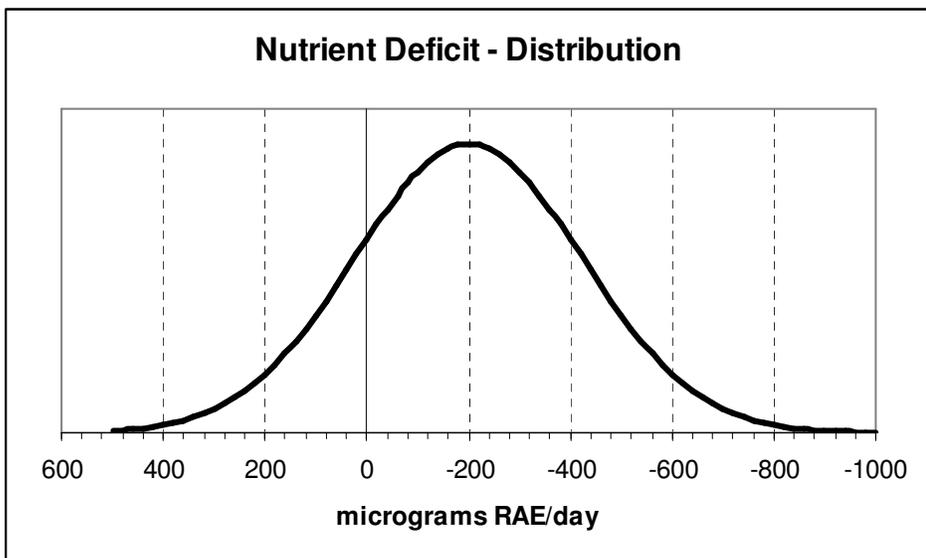
Distribution of the adult female requirement for Vitamin A

Normal with mean = 500 and SD = 100

Distribution of distribution of Vitamin A in a hypothetical diet

Normal with mean = 700 and SD = 200

FIGURE 5b



Distribution of Requirement minus Intake (Nutrient Deficit) of Vitamin A for an adult female on a hypothetical diet

Normal with mean = -200 and SD = 224

The area for which the Nutrient Deficit is positive (Intake below Requirement) is 18.6%; this is the risk of Vitamin A deficiency in an adult female on this diet.