American Journal of Epidemiology
Vol. 137, No. 12
Printed in U.S.A.

A BRIEF ORIGINAL CONTRIBUTION

Interpretation of Energy Adjustment Models for Nutritional Epidemiology

Victor Kipnis,[1] Laurence S. Freedman,[1] Charles C. Brown,[1] Anne Hartman,[2]
Arthur Schatzkin,[3] and Sholom Wacholder[4]

The authors discuss the interpretation of three alternative energy adjustment models for nutritional epidemiology. It is shown that four different effects are addressed by these models: 1) adding nutrient N, 2) substituting nutrient N for "other" nutrients, 3) adding "other" nutrients, and 4) adding both N and "other" nutrients in a specific ratio. Each of these effects may be estimated from any of the three models. The relative standard errors for the four estimated effects are also provided. *Am J Epidemiol* 1993; 137:1376–80.

epidemiologic methods; models, statistical; nutrition

With regard to the paper by Willett and Stampfer (1) and subsequent correspondence from Pike et al. (2) and Howe (3), we wish to point out some further aspects of the various proposed statistical models for analyzing relations between a disease and certain macronutrient intakes.

THE MODELS AND THEIR INTERPRETATION

The three sets of authors mentioned above have considered the following models: the Standard Model (4),

$$D = \beta_{0S} + \beta_{1S}N + \beta_{2S}T + \epsilon,$$

the Residual Model (1),

$$D = \beta_{0R} + \beta_{1R}R + \beta_{2R}T + \epsilon,$$

and the Energy Partition Model (3),

$$D = \beta_{0P} + \beta_{1P}N + \beta_{2P}(T - N) + \epsilon,$$

where D is the disease indicator variable, T is the total energy intake, N is the intake of calories from nutrient N, and ϵ is random error. We assume that the variables T and N are measured in kilocalories per day. The variable R is the "energy-adjusted nutrient intake," that is, the residual from regressing N on T (1). Specifically,

$$R = N - \alpha_0 - \alpha_1 T,$$

where α_0 and α_1 are the regression coefficients in the linear regression of N on T. We assume here for simplicity that D is a continuous variable and the above equations are linear regression models. However, the ensuing remarks can be applied, at least qualitatively, to logistic regression. We do not consider in this paper the Nutrient Density Model (4) or the case of micronutrients. Both of these topics deserve separate consideration.

As Pike et al. (2) and Howe (3) have

Received for publication August 11, 1992, and in final form February 8, 1993.
[1] Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD.
[2] Applied Research Branch, Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD.
[3] Cancer Prevention Studies Branch, Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD.
[4] Biostatistics Branch, Division of Cancer Etiology, National Cancer Institute, Bethesda, MD.
Reprint requests to Dr. Victor Kipnis, Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute, Executive Plaza North, Suite 344, Bethesda, MD 20892.

already pointed out, these models are mathematically equivalent and could be viewed as reparameterizations of the same model. However, each model, when used in the routine way to develop estimates and significance tests of the coefficients, addresses different questions. In multiple regression models such as those above, the coefficient attached to one variable (e.g., β_{1S}) can be interpreted as the effect upon disease of increasing that variable by one unit while keeping the other variable constant. For example, β_{1S} represents the effect of increasing nutrient N by 1 kcal while keeping total calories unchanged. This can only be achieved by a simultaneous reduction of "other" nutrients by 1 kcal. Thus, β_{1S} represents the effect of substituting calories from nutrient N for "other" calories while keeping total calories constant. Similarly, in the Standard Model, β_{2S} represents the effect of increasing total energy intake by 1 kcal while keeping energy intake from nutrient N constant. This is equivalent to increasing calories from "other" nutrients by 1 kcal. Thus, β_{2S} represents the effect of increasing "other" nutrients.

In the Partition Model, β_{1P} represents the effect of increasing total calories by increasing calories from nutrient N while keeping "other" calories constant. Clearly, the interpretation of β_{1P}, the effect of adding nutrient N, is different from the interpretation of β_{1S}, the effect of substituting nutrient N for "other" nutrients. The coefficient β_{2P} represents the effect of an overall increase of 1 kcal in total calories by increasing "other" calories while keeping nutrient N calories constant—that is, the same effect as β_{2S}.

Similar considerations make it clear that in the Residual Model, β_{1R} represents the same effect as β_{1S}, that is, the effect of substituting calories from nutrient N for "other" calories. The coefficient β_{2R} represents the effect on disease of increasing total energy by 1 kcal while keeping nutrient residual R constant. Since $R = N - \alpha_0 - \alpha_1 T$, when total energy T is increased by 1 kcal, residual R is kept constant only by increasing the intake of nutrient N by α_1 kcal and the intake of other nutrients by $(1 - \alpha_1)$ kcal. Thus, the coefficient β_{2R} describes the effect of increasing the intake of a diet comprising calories from nutrient N and "other" nutrients in the specific proportions α_1 and $1 - \alpha_1$, respectively.

One may summarize these points by noting that four different effects are addressed by these models, namely 1) adding nutrient N while keeping "other" nutrients constant, 2) substituting nutrient N for "other" nutrients, 3) adding "other" nutrients while keeping nutrient N constant, and 4) adding both N and "other" nutrients in a specific ratio. Each of these effects may be estimated from any of the three models.

The effect of adding nutrient N is estimated by

$$\beta_{1P} \text{ or } \beta_{1S} + \beta_{2S} \text{ or}$$
$$(1 - \alpha_1)\beta_{1R} + \beta_{2R}.$$

The effect of substituting nutrient N for "other" nutrients is

$$\beta_{1S} \text{ or } \beta_{1R} \text{ or } \beta_{1P} - \beta_{2P}.$$

The effect of adding "other" nutrients is

$$\beta_{2S} \text{ or } \beta_{2P} \text{ or } \beta_{2R} - \alpha_1\beta_{1R},$$

and the effect of adding both N and "other" nutrients in the proportions α_1 and $1 - \alpha_1$ is

$$\beta_{2R} \text{ or } \alpha_1\beta_{1P} + (1 - \alpha_1)\beta_{2P}$$
$$\text{or } \alpha_1\beta_{1S} + \beta_{2S}.$$

The above equivalences can be obtained algebraically from a comparison of the three models using the equations linking T, N, and R, as described by Pike et al. (2), among others. Some of the equivalences can be easily understood from the interpretations given above. For example, β_{1S} or β_{1R}, the effect of substituting N for "other" nutrients, equals the difference between β_{1P}, the effect of adding nutrients N, and β_{2P}, the effect of adding "other" nutrients.

STATISTICAL PRECISION

There has been little discussion in the literature so far of the standard errors of the estimated coefficients. Howe (3) does mention that the standard errors of the coefficients from one model may be used directly to calculate the standard errors for the other models. We note first that the standard errors of any of the equivalent estimators, given above for each of the four effects, are equal. For example, the standard errors of the estimators of β_{1P} or $\beta_{1S} + \beta_{2S}$ or $(1 - \alpha_1)\beta_{1R} + \beta_{2R}$ are all equal to each other, and the same applies to any of the other sets of equivalent estimators. Second, as is shown in the Appendix, the precision of evaluating the four effects mentioned above varies. In fact, the standard errors for the four estimated effects, taken in the same order as listed above, are in the ratios

$$1:\sqrt{k^2 + 2k\rho + 1}/k:1/k:$$
$$\sqrt{(1 - \rho^2)/(k^2 + 2k\rho + 1)},$$

where ρ is the correlation between "other" nutrients and nutrient N calories and k is the ratio of their standard deviations. In most applications, the correlation between nutrient N and "other" nutrients will be positive, and in this case the smallest standard error will be for the fourth effect, i.e., adding both N and "other" nutrients in a specific proportion. Furthermore, the estimate of the second effect, i.e., substituting nutrient N for "other" nutrients, will be the least precise of the four. The precision of estimating the effect of adding nutrient N (the first effect) is greater than that of estimating the effect of adding "other" nutrients (the third effect) if $k < 1$, and it is smaller if $k > 1$. To provide an idea of the relative precision for estimating these four effects, we use data on fat and nonfat intake from the Women's Health Trial Vanguard Study (V. Kipnis, unpublished data). We estimated ρ to be 0.53 and k to be 1.66. These data would lead to standard errors for the four effects in the ratio 1:1.4:0.6:0.4. Thus, for example, we can estimate the effect of *adding* fat (the first effect) 1.4 times more precisely than the effect of *substituting* fat for "other" nutrients (the second effect).

SUMMARY

Our main message in this paper is that the primary decision of the investigator is to decide which of the four effects described above are of interest. Once this is done, any of the three statistical models may be used to estimate and test the effect of interest. However, it is much simpler to employ a model in which the effect of interest is represented by a single regression coefficient. For example, if one were mainly interested in the effect of substituting nutrient N for "other" nutrients, the Standard Model or the Residual Model would provide a natural estimate for the effect through evaluation of β_{2S} or β_{1R}, and its standard error would routinely be calculated as part of the analysis. If the Partition Model were used, the effect could still be easily evaluated through $\beta_{1P} - \beta_{2P}$, but the standard error of this estimate would not routinely be provided in the analysis, and its calculation would require knowledge of the covariance of the estimated regression coefficients. Thus, the "natural" models are: for the effect of adding nutrient N, the Partition Model; for the effect of substituting nutrient N, the Standard Model or the Residual Model; for the effect of adding "other" nutrients, the Standard Model or the Partition Model; and for the effect of adding both N and "other" nutrients (in the specific proportion mentioned), the Residual Model.

REFERENCES

1. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986;124:17–27.
2. Pike MC, Bernstein L, Peters R. Re: "Total energy intake: implications for epidemiologic analyses." (Letter). Am J Epidemiol 1989;129:1312–13.
3. Howe GR. The first author replies. (Reply to letter). Am J Epidemiol 1989;129:1314–15.
4. Willett W. Nutritional epidemiology. New York: Oxford University Press, 1990:245–71.

APPENDIX

In a multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon,$$

with response variable y and two stochastic explanatory variables x_1 and x_2, the asymptotic variance of the estimated regression coefficient $\hat{\beta}_i$, $i = 1, 2$, is given as

$$\operatorname{var}(\hat{\beta}_i) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_i^2\,(1 - \rho^2(x_1, x_2))}\right), \tag{A1}$$

where $\sigma_i^2 = \operatorname{var}(x_i)$, $i = 1, 2$; $\sigma^2 = \operatorname{var}(\epsilon)$; $\rho(x_1, x_2)$ is the correlation coefficient between x_1 and x_2; and n is the sample size. Let ρ denote the correlation between "other" nutrients and nutrient N calories, and k the ratio of their standard deviations. Since $T = (T - N) + N$, the variance of σ_T^2 of the total caloric intake is calculated as

$$\sigma_T^2 = \sigma_{T-N}^2 + 2\rho\sigma_{T-N}\sigma_N + \sigma_N^2 = \sigma_N^2(k^2 + 2k\rho + 1), \tag{A2}$$

where σ_{T-N}^2 and σ_N^2 are the variances of the "other" nutrients and nutrient N calories, respectively. The covariance between variables N and T is

$$\operatorname{cov}(N, T) = \operatorname{cov}(N, T - N) + \operatorname{var}(N) = \rho\sigma_N\sigma_{T-N} + \sigma_N^2 = \sigma_N^2(k\rho + 1),$$

and their correlation coefficient is given by

$$\rho_{N,T} = \frac{\operatorname{cov}(N, T)}{\sigma_N\sigma_T} = \frac{k\rho + 1}{\sqrt{k^2 + 2k\rho + 1}}. \tag{A3}$$

Because of its definition, the residual R does not correlate with total calorie intake T,

$$\rho(R, T) = 0, \tag{A4}$$

and its variance σ_R^2 is given by $\sigma_R^2 = \sigma_N^2\,(1 - \rho_{N,T}^2)$, which according to equation A3 can be rewritten as

$$\sigma_R^2 = \sigma_N^2\,\frac{k^2(1 - \rho^2)}{k^2 + 2k\rho + 1}. \tag{A5}$$

Substituting equations A2–A5 into equation A1 for each of the three considered energy adjustment models provides the following asymptotic variances of the corresponding regression coefficients:

1) Standard Model:

$$\operatorname{var}(\hat{\beta}_{1S}) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_N^2(1 - \rho_{N,T}^2)}\right) = \frac{\sigma^2}{n}\left(\frac{k^2 + 2k\rho + 1}{\sigma_N^2 k^2(1 - \rho^2)}\right),$$

$$\operatorname{var}(\hat{\beta}_{2S}) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_T^2(1 - \rho_{N,T}^2)}\right) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_N^2 k^2(1 - \rho^2)}\right);$$

2) Residual Model:

$$\operatorname{var}(\hat{\beta}_{1R}) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_R^2(1 - \rho_{R,T}^2)}\right) = \frac{\sigma^2}{n}\left(\frac{k^2 + 2k\rho + 1}{\sigma_N^2 k^2(1 - \rho^2)}\right),$$

$$\operatorname{var}(\hat{\beta}_{2R}) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_T^2(1 - \rho_{R,T}^2)}\right) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_N^2(k^2 + 2k\rho + 1)}\right);$$

3) Energy Partition Model:

$$\mathrm{var}(\hat{\beta}_{1P}) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_N^2(1-\rho^2)}\right),$$

$$\mathrm{var}(\hat{\beta}_{2P}) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_{T-N}^2(1-\rho^2)}\right) = \frac{\sigma^2}{n}\left(\frac{1}{\sigma_N^2 k^2(1-\rho^2)}\right).$$

The standard errors for the four estimated effects listed in the text are therefore given by the following expressions.

The standard error (SE) of the effect of adding nutrient N is

$$\mathrm{SE}(\hat{\beta}_{1P}) = \frac{\sigma}{\sigma_N\sqrt{n}}\left(\frac{1}{\sqrt{1-\rho^2}}\right).$$

The standard error of the effect of substituting nutrient N for "other" nutrients is

$$\mathrm{SE}(\hat{\beta}_{1S}) = \mathrm{SE}(\hat{\beta}_{1R}) = \frac{\sigma}{\sigma_N\sqrt{n}}\left(\frac{\sqrt{k^2+2k\rho+1}}{k\sqrt{1-\rho^2}}\right).$$

The standard error of the effect of adding "other" nutrients is

$$\mathrm{SE}(\hat{\beta}_{2S}) = \mathrm{SE}(\hat{\beta}_{2P}) = \frac{\sigma}{\sigma_N\sqrt{n}}\left(\frac{1}{k\sqrt{1-\rho^2}}\right).$$

The standard error of the effect of adding both N and "other" nutrients in the proportions α_1 and $1-\alpha_1$ is

$$\mathrm{SE}(\hat{\beta}_{2R}) = \frac{\sigma}{\sigma_N\sqrt{n}}\left(\frac{1}{\sqrt{k^2+2k\rho+1}}\right).$$