Within- and Between-Cohort Variation in Measured Macronutrient Intakes, Taking Account of Measurement Errors, in the European Prospective Investigation into Cancer and Nutrition Study

Pietro Ferrari1, Rudolf Kaaks2, Michael T. Fahey1, Nadia Slimani1, Nicholas E. Day3, Guillem Pera4, Hendrik C. Boshuizen5, Andrew Roddam6, Heiner Boeing7, Gabriele Nagel8, Anne Thiebaut9, Philippos Orfanos10, Vittorio Krogh11, Tonje Braaten12, and Elio Riboli1

1 Unit of Nutrition and Cancer, International Agency for Research on Cancer, Lyon, France. 2 Hormones and Cancer Group, International Agency for Research on Cancer, Lyon, France. 3 Institute of Public Health, University of Cambridge, Strangeways Research Laboratory, Cambridge, United Kingdom. 4 SERC, Institut Català d’Oncologia, Barcelona, Spain. 5 Centre for Information Technology and Methodology, National Institute for Public Health and the Environment, Bilthoven, the Netherlands. 6 Cancer Research UK Epidemiology Unit, Oxford University, Oxford, United Kingdom. 7 Division of Clinical Epidemiology, German Cancer Research Centre, Heidelberg, Germany. 8 German Institute of Human Nutrition, Potsdam-Rehbrücke, Germany. 9 E3N-EPIC, INSERM Z521, Institut Gustave Roussy, Villejuif, France. 10 Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece. 11 Epidemiology Unit, Istituto Nazionale dei Tumori, Milan, Italy. 12 Institute of Community Medicine, University of Tromsø, Tromsø, Norway.

Received for publication October 2, 2003; accepted for publication April 8, 2004.

Multicenter epidemiologic studies provide a unique opportunity to evaluate the association between exposure and disease at the individual and the aggregate levels. The two components can eventually be pooled to corroborate each other, using weights proportional to the intraclass correlation coefficient (ICC), which expresses the amount of between-cohort variability in the exposure variable compared with the total. The greater the ICC, the more the overall estimate will reflect the between-cohort component. Dietary measurements are affected by measurement errors, particularly within a cohort. In 1992–2000, the variability of macronutrient intake distribution before and after calibration for measurement error in the European Prospective Investigation into Cancer and Nutrition was evaluated. A two-level, random-effects model was used. Evaluation of macronutrient densities revealed that energy has a considerable effect on the calibration model, leading to ICC values larger than those for the absolute intakes. Given the shrinkage of the within-center variability, a sizable increase in the ICC was observed for protein in men and women (0.48 and 0.54, respectively) and carbohydrates in men (0.41). Results suggest that the effect of calibration on macronutrient intake variability is greater for the within-cohort component, thus increasing the relative importance of the between-cohort component. After calibration, the two components had a similar weight. This observation has important implications for the analysis of multicenter studies because the between-cohort component provides a large part of the overall heterogeneity.

calibration; diet; epidemiologic methods; multicenter studies

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; ICC, intraclass correlation coefficient.

Ecologic correlation studies (i.e., those based on international comparisons) have shown strong associations between cancer risk and average per capita consumption of specific food groups and nutrients (1). This initial observation—complemented by data from time-trend studies and migrant studies, which confirm the importance of environmental and lifestyle factors in determining cancer risk—has led to a vast research effort to identify diet-cancer relations at the level of single individuals (2). After hundreds of case-control studies, this effort has gradually shifted toward those of a

Correspondence to Dr. Pietro Ferrari, IARC, 150, cours Albert-Thomas, 69372 Lyon Cedex 08, France (e-mail: Ferrari@iarc.fr).
Within- and Between-Cohort Macronutrient Variation

prospective cohort design because of concern that case-control studies may be biased by selection of improper controls or by differential errors between cases and controls in their estimates of habitual diet.

Compared with ecologic studies, investigations that use individual-level information on dietary exposures and disease outcome have the advantage of allowing more complex analyses, with adjustments for potential confounding factors and evaluations of possible interactions between multiple exposures. However, a possible limitation is that random errors in assessing individuals’ habitual dietary intake can be so large as to virtually mask potentially existing associations with disease risk (3, 4). The impact of random measurement error on estimated measures of diet-disease associations depends on the variance of errors relative to the between-person variance of true intake levels (5).

For practical and logistic reasons, the increase in statistical power that may be obtained by increasing study size or reducing the average magnitude of random dietary assessment errors will generally be limited. However, an additional possibility for improving the statistical power is to increase the true overall heterogeneity of the dietary exposures studied, combining data from multiple studies conducted in populations with different dietary habits. This rationale was used to set up the European Prospective Investigation into Cancer and Nutrition (EPIC), a multicenter cohort study on diet and cancer conducted in 28 regional centers nested in 10 Western European countries with widely varying dietary practices and cancer risks (6).

In a multicohort study, the overall association between diet and cancer risk can be broken down into 1) within-cohort relations, which reflect the associations at the individual level in each of the cohorts; and 2) a between-cohort relation, which captures the association between exposure and disease risk at the aggregate level. For proper comparison, it is crucial that corrections be made for any between-cohort differences that may exist in systematic exposure measurement errors in reporting individuals’ diet (i.e., totally due to within-person random variability or to random measurement errors in reporting individuals’ diet). This observation may have important implications for the analysis of the EPIC study as well as for the design of other prospective studies.

MATERIALS AND METHODS

EPIC data collection

EPIC is a multicenter cohort study aimed at investigating the association between diet and cancer across Western Europe. Individual habitual dietary intake was assessed by using different questionnaires developed and administered independently in each country. To express individual dietary intakes according to the same reference scale and to correct the diet-disease associations for attenuation due to measurement errors, a single 24-hour dietary recall was collected by using a software program (EPIC-SOFT) specially designed to standardize the recall interviews (12). The 24-hour dietary recalls are used as reference measurements and were collected from a stratified sample of about 37,000 men and women in the EPIC cohort (13). For the present analyses, study subjects were recruited in 1992–2000.

Calibration model

We assumed the following measurement error model for the baseline dietary assessment measurements (Q) and for the 24-hour dietary recall measurement (R) in relation to the unknown true intake (T):

\[
Q = \alpha_Q + \beta_Q T + \varepsilon_Q \tag{1a}
\]

\[
R = T + \varepsilon_R, \tag{1b}
\]

where \(E(\varepsilon_Q | T) = 0, E(\varepsilon_R | T) = 0\), \(\text{var}(\varepsilon_Q) = \sigma_{\varepsilon_Q}^2\), and \(\text{var}(\varepsilon_R) = \sigma_{\varepsilon_R}^2\). The association shown in model 1b—for reference measurement R—assumes that errors are strictly random and that variation around individuals’ unknown true intake is totally due to within-person random variability or to random measurement errors in reporting individuals’ diet (i.e., \(\text{cov}(\varepsilon_Q, T) = 0\)). For the questionnaire measurements Q, the coefficients \(\alpha_Q\) and \(\beta_Q\) in model 1a express constant and proportional scaling biases (14), whereas the residual term \(\varepsilon_Q\) models the random part of measurement errors in Q and is assumed to be uncorrelated with true intake level T (15).

To extend the calibration methodology introduced by Rosner et al. (10), a very general form of the EPIC calibration model can be defined as

\[
R_{ij} = \alpha_j + \lambda_W(Q_{uj} - \bar{Q}_j) + \lambda_B \bar{Q}_j + \gamma_W(Z_{uj} - \bar{Z}_j) + \gamma_B \bar{Z}_j + \varepsilon_{ij}, \tag{2}
\]

where
\[
\begin{align*}
\alpha_j &= \alpha + u_{0j}, \quad \lambda_{Wj} = \lambda_W + u_{1j} \\
\left( u_{0j}, u_{1j} \right) &\sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{0j} & \sigma_{01}^2 \\ \sigma_{01}^2 & \sigma^2_{1j} \end{pmatrix} \right) \quad \text{and} \quad \epsilon_{ij} \sim N(0, \sigma^2_{\epsilon}),
\end{align*}
\]

where \( j = 1 \ldots G \) indexes the group variable, be it the recruitment region, center, or country; \( i = 1 \ldots n_j \) are individual measurements within group \( j \); and \( \bar{Q}_j \) are the group means.

The coefficients \( \gamma_j \)’s model the effect of those variables \((Z)\) included as confounding factors, for example, age, body mass index, physical activity, and smoking status. Models 1a and 1b and model 2 assume independence between random errors in \( R \) and \( Q \), which is unlikely to occur in self-reported measurements as a consequence of study subjects’ tendency to consistently under- or overestimate dietary intake \((14, 16–19)\). However, in the absence of a third objective measure of intake, such as a biomarker, the assumption of independence between the errors in \( R \) and \( Q \) cannot be assessed. One single replicate of \( R \) allows a correct estimate of the attenuation factors \((11)\) to be obtained in model 2. This model has fixed-effect parameters \((\alpha, \lambda_j, s)\) and \( \gamma_j \)’s and three random effects, \( u_{0j}, u_{1j} \), and \( \epsilon_{ij} \), which give rise to four variance components reflecting, respectively, variation between groups \((\sigma^2_{0j})\), variability in the calibration coefficients across groups \((\sigma^2_{\epsilon})\), their covariance \((\sigma_{0\epsilon})\), and within-group variation among study subjects \((\sigma^2_{\epsilon})\). In model 2, \( \lambda_{Wj} \) captures the within-group component of the calibration model, while \( \lambda_j \) reflects the between-group component. Multilevel models provide a unique opportunity to simultaneously model within- and between-group effects \((20–22)\).

Use of a random-effects model allows the heterogeneity of attenuation factors to be estimated parsimoniously. It also enables evaluation of the agreement between \( R \) and \( Q \) measurements at the aggregate level in model 2 \((20, 23)\).

Within center, the calibration factor \( \lambda_{Wj} \) can be used to correct the observed ("naïve") parameter that estimates the association between dietary exposure and the outcome of interest, for example, cancer incidence, by

\[ \hat{\beta}_{\text{True}(j)} = \hat{\beta}_{\text{Obs}(j)}/\hat{\lambda}_{Wj}, \]

where \( \hat{\beta}_{\text{Obs}(j)} \) is the naïve estimate that expresses the increase in risk for a unit increase in the exposure and is obtained by relating cancer incidence to the \( Q \) measurements, following a "disease model." A mathematically equivalent way to correct for measurement error in the diet/disease association is to estimate \( E(T|Q) \) as the predicted values of the calibration model and use these predicted values as the imputed exposures in the disease model.

### Variability of exposure in the EPIC study

To evaluate the effect of calibration in a multicenter setting, we focused on the distribution of predicted exposure values, \( E(T|Q) \). As a result of the calibration model considered, the variance of predicted within-group values, compared with \( Q \), will shrink by a factor proportional to \( \lambda_{Wj}^2 \) \((11)\). Using the calibration sample in the EPIC study \((13)\), we calculated the predicted values according to calibration model 2. To estimate the between- and within-group variance components in the "calibrated" exposure, a mixed model was used. Here, we refer to this model as the "evaluation" model, where the observed \( Q \) and \( R \) measurements and the predicted intake values are each used as response variables in separate analyses, and the variance components are estimated by modeling the group effect with a random effect as

\[
Y_{ij} = \alpha_{Ej} + \gamma_{EjW}Z_{ij} - Z_j + \gamma_{EjR}W_j + \epsilon_{Eij} \\
\alpha_{Ej} = \alpha_E + u_{Eij},
\]

where \( Y_{ij} \) being either \( Q_{ij}, R_{ij}, \) or \( E(T|Q) \) estimated in model 2. This model enables estimation of the proportion of total variation in the exposure attributable to the between-group component, expressed by intraclass correlation coefficient \((\text{ICC}) = \sigma^2_1/(\sigma^2_1 + \sigma^2_2)\). By symmetry, \( 1 - \text{ICC} \) expresses the variability attributable to the within-group component.

### Energy and macronutrient variability

We considered statistical models for intakes of protein, lipids, carbohydrate, alcohol, and total energy. In the absence of a standardized European nutrient database, country-specific food composition tables were used to calculate nutrient intakes for both \( Q \) and \( R \). Results from a recent validation study \((24)\) showed higher correlations between dietary protein from \( R \) measurements and protein estimated from urinary nitrogen, compared with \( Q \). This finding suggests that the dietary assessment method used probably has a higher impact on measurement errors than use of local nutrient databases.

To adjust for total energy, macronutrient densities \((\text{i.e., macronutrient intakes divided by total energy intake})\) were also considered. ICC estimates were computed for nutrient intake in \( Q \) and \( R \) measurements, as well as for predicted values, according to model 2 for absolute intakes and density values.

To model the group effect in the present analysis, 28 centers were considered. The macronutrient distributions approximated a normal distribution, whereas the distribution of alcohol intake was rather skewed. However, the present analyses focused primarily on point estimates of fixed effects and variance components. These quantities are fairly robust to departures from normality in mixed models \((25)\). Results using log-transformed values were very similar to those obtained by using untransformed variables.

The calibration and the evaluation models were all adjusted for age and body mass index at the individual and at the group levels. These variables were chosen because previous studies \((26–29)\) have suggested that they are significant determinants of the accuracy of \( Q \) measurements. To allow the effect of age and body mass index to be heterogeneous across centers, random effects for these variables were included in model 2, but the associated variance components were not significantly different from zero, thus suggesting

---

homogeneity of effects. Overall, linear fixed effects for age and body mass index, compared with models with higher-order terms, best fit the data.

To explore the role of confounding factors in explaining the heterogeneity of the attenuation factors, we also considered a calibration model with a cross-level interaction between Q measurements and mean values of age and body mass index. Doing so enables exploration of the dependency of attenuation factors on predictors.

Gender-specific analyses were performed. Statistical analyses were carried out by using the MIXED procedure of SAS software (30). Parameter estimates were obtained by restricted maximum likelihood estimation. A “sandwich” estimator was used to compute corrected standard errors involving fixed-effects parameters, which are robust against model misspecification (25, 31, 32) in model 2. The likelihood ratio statistic based on restricted maximum likelihood estimation was used to test the significance of random effects.

RESULTS

In tables 1 and 2, the square root of the between- and within-center variance components and corresponding ICC estimates for energy intake and the macronutrient distributions of Q, R, and E(T|Q) measurements are reported for men and women, respectively. These tables also show estimates for the between- and within-center components of the calibration model, \( \lambda_B \) and \( \lambda_W \), according to model 2. Results for macronutrients are reported with and without adjustment for energy intake.

For men and women, respectively, the calibration factors for the within-center component ranged from 0.36 and 0.31 for protein intake to 0.79 and 0.84 for alcohol. Similar values for the between-center \( \lambda_B \) were obtained, ranging from 0.03 for lipids in men and 0.10 for energy intake in women to 1.04 and 1.29 for alcohol in men and women. Adjustment for energy intake had very little impact on \( \lambda_W \), whereas between-center calibration factors were higher regarding energy densities for carbohydrates in men (\( \lambda_B \) from 0.65 to 0.90) and...
women ($\lambda_B$ from 0.41 to 0.76) and for lipids in men ($\lambda_B$ from 0.03 to 0.85) and women ($\lambda_B$ from 0.33 to 0.75). Lower values were observed for protein in men ($\lambda_B$ from 0.44 to 0.17).

The square root of the variance component of $\lambda_W$, $\sigma^2$, which estimates the variability of attenuation factors across EPIC centers, was consistently and significantly different from zero for both men and women. This finding suggests that the accuracy of $Q$ was rather heterogeneous across the EPIC centers.

Variance components for $Q$ measurements showed very similar patterns for between-center variability in men and women. Respectively, we found low ICC values for alcohol (0.11 and 0.06), moderately low ICCs for energy intake (0.14 and 0.13) and lipids (0.10 and 0.15), and slightly higher ICCs for protein (0.25 and 0.21) and carbohydrates (0.17 for both genders).

The patterns of ICC values for $R$ measurements were similar, but the values were consistently lower than 0.10. The between-center variance components were overall lower in $R$ than in $Q$, except regarding lipids in men. On the other hand, within-center variation was considerably higher in $R$, a result that very likely reflected day-to-day variability in the single dietary measurements available for each study participant in the calibration subsample. For energy intake, within-center variability was 1.8 and 1.6 times larger in $R$ than in $Q$ measurements for men and women, respectively. For men and women, the same ratios were 2.0 and 1.5 for protein, 1.6 and 1.4 for carbohydrates, 2.5 and 2.1 for lipids, and 2.1 and 2.4 for alcohol.

For the predicted intake $E(T|Q)$, the between-center variability was very close, but not identical, to the variation in the $R$ measurements as a consequence of including a random effect ($u_{ij}$) for the center variable in model 2. The within-center variability of $E(T|Q)$ shrank when compared with $Q$ measurements.

Adjustment for energy intake had a variable impact on the estimated variability of macronutrients. For alcohol and

### Table 2: Estimated calibration coefficients and between- and within-center variance components of absolute values and densities of macronutrient intake for women in the European Prospective Investigation into Cancer and Nutrition study, 1992–2000*

<table>
<thead>
<tr>
<th>Exposure and type†</th>
<th>Absolute intake value</th>
<th>Density value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_B$‡</td>
<td>$\lambda_W$§</td>
</tr>
<tr>
<td>Energy intake</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>194</td>
<td>494</td>
</tr>
<tr>
<td>$R$</td>
<td>98</td>
<td>617</td>
</tr>
<tr>
<td>$E(T</td>
<td>Q)$</td>
<td>0.10</td>
</tr>
<tr>
<td>Protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>$R$</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>$E(T</td>
<td>Q)$</td>
<td>0.44</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>30</td>
<td>65</td>
</tr>
<tr>
<td>$R$</td>
<td>16</td>
<td>77</td>
</tr>
<tr>
<td>$E(T</td>
<td>Q)$</td>
<td>0.41</td>
</tr>
<tr>
<td>Lipids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>$R$</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>$E(T</td>
<td>Q)$</td>
<td>0.33</td>
</tr>
<tr>
<td>Alcohol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>$R$</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>$E(T</td>
<td>Q)$</td>
<td>1.29</td>
</tr>
</tbody>
</table>

* Results were adjusted for age and body mass index.
† $Q$, dietary questionnaire measurements; $R$, 24-hour dietary recall measurements; $E(T|Q)$, predicted values.
‡ Between-center calibration factor in model 2. Refer to the Materials and Methods section of the text for an explanation of the models.
§ Within-center calibration factor in model 2.
¶ Square root of the variance component for the within-center calibration factor ($\lambda_W$) in model 2, consistently significantly different from zero, according to likelihood ratio test.
# Square root of the variance component for the between-center component in evaluation model 3.
** Square root of the variance component for the within-center component in evaluation model 3.
†† ICC, intraclass correlation coefficient.
protein in men and women, ICC values for predicted intakes were basically unaffected. Adjustment for energy had a moderate effect for carbohydrates in men (ICC from 0.41 to 0.52). Decidedly higher differences in ICC values were observed for macronutrient densities regarding carbohydrates in women (ICC from 0.26 to 0.51) and lipids in men (ICC from 0.35 to 0.61) and women (ICC from 0.39 to 0.60). Very little change in the variance components for Q and R measurements was observed for macronutrient density values.

For both absolute intake and density values, ICC values overall were between 1.5- and 3.0-fold higher for $E(T|Q)$ than for $Q$, thus indicating strong shrinkage in the within-center variance component after calibration. For lipid density in men, the ICC ranged from 0.20 for $Q$ measurements to 0.61 for predicted values.

Overall, negative effects were observed in calibration model 2 for age and body mass index. For alcohol intake in men, the effects of body mass index were positive (data not shown). Results from the model with a cross-level interaction to explore the heterogeneity of attenuation factors revealed overall nonsignificant associations. However, in the case of carbohydrate and lipid density, of alcohol intake in men, and of protein density in women, a significant negative interaction with body mass index was observed, thus suggesting that centers in which mean body mass index levels were higher tended to have lower $Q$ accuracy (data not shown).

**DISCUSSION**

In multicenter cohort studies, associations of exposure with disease risk can be studied at two complementary levels: within and between cohorts (countries or recruitment centers). Proper comparability of both the within- and between-center relations, and proper relative weighting of these components of the overall association, requires a high degree of standardization of exposure measurements as well as measured disease outcomes (7). Furthermore, corrections must be made for variable degrees of random measurement error that, within center, will generally attenuate relative risk estimates.

In the context of the European EPIC study on diet and cancer, we have described the use of calibration studies to correct for between-cohort differences in effects of dietary assessment errors, putting the measurements on a more standardized scale. The calibration was based on collection, within a large, stratified subsample of the EPIC cohorts, of standardized R measurements. Regression of these "reference" measurements on the individual-level questionnaire data enabled estimation of predicted values $E(T|Q)$, indicating average true intake levels for given $Q$ measurements that, in a second phase, can be used as predictors in a disease model. A two-level, random-effects model was used to estimate within- and between-center calibration effects. The calibration model included adjustments for effects of potential confounders, particularly total energy intake, body mass index, and age.

Random effects allowed parsimonious modeling of the heterogeneity of effects across EPIC centers without having to include a large number of interaction terms. Random-effects models also allow the inclusion of aggregate (contextual) variables to be evaluated and the exploration of possible sources of between-center differences in dietary questionnaire accuracy, by making such differences dependent on aggregate-level predictors.

We decided to focus on the effects of calibration on within- and between-center variance components of macronutrient intake levels. The macronutrient composition of diet is still a major factor of interest in the nutritional epidemiology of cancer. In a relatively recent review (33), the macronutrients were considered comparable (e.g., in the case of alcohol) or readily comparable after correction for different methods of calculation and mode of expression (as for energy, protein, and carbohydrates). In addition, a validation study by Slimani et al. (24), recently conducted in the EPIC cohort, supported this finding, where relatively high correlations were observed in an ecologic comparison of protein intake estimated from urinary nitrogen and dietary protein, estimated by $R$ and $Q$ measurements. A more extensive, fully standardized European food composition database, also comprising all detailed micronutrients, is currently being prepared for the EPIC study (34).

The lack of common food composition tables might lead to a larger between-center variability, as a consequence of systematic food composition tables-specific errors, randomly distributed across countries, but it is less likely to affect the sizable change in ICC values observed for predicted intakes. In the present analyses, the calibration model took into account the geographic specificity of the EPIC entities, by performing center-specific comparisons between $R$ and $Q$ measurements, for which the same national nutrient databases were used. After correction for measurement error, we believe that the shrinkage of within-center macronutrient variability rather than the inflation of between-center variability was responsible for the relative increase in ICC values for $E(T|Q)$ compared with $Q$. As expected, between EPIC centers, mean values of $E(T|Q)$ were practically identical to mean levels of $R$. The term $\lambda_y$ provides an estimate of the between-center (linear) relations between mean $R$ and $Q$ measurements. Rather low values of the coefficient were observed for lipids in men and total energy intake in women, but values approximating 1 were observed for alcohol, particularly in men, thus suggesting very good agreement between $R$ and $Q$ at the aggregate level. Because special efforts have been made to standardize the $R$ measurements to estimate absolute, mean dietary intake levels, we assume that the mean calibrated values reflect more accurately the between-cohort variations in mean true intake than the mean $Q$ values do (23).

Within EPIC centers, the distributions of predicted intake values (i.e., after calibration) generally showed strong shrinkage toward center-specific mean values. The shrinkage in within-center variance was proportional to the square of within-center calibration coefficients $\lambda_{yj}$. The shrinkage depended only very moderately on the effects of other individual-level variables used to explain $R$ variability, such as age and body mass index.

The association between exposure and disease can be decomposed as $\beta_{\text{Total}} = \beta_{\text{ICC}}\text{ICC} + \beta_{\text{W}}(1 - \text{ICC})$, where the $\beta$'s...
express the ecologic (B) and individual (W) association between exposure and disease, provided that the within-centers relations are homogeneous. After calibration, at least for macronutrients, a sizable weight is given to the between-cohort component of the multicenter EPIC analysis, and the weights of the two levels of evidence of the diet/disease association are similar.

Results from the present analyses suggest that the between-center variability of exposure provides a sizable amount of information and should therefore be exploited. Analysis of the diet/disease association in the EPIC study should provide estimates of the within- and between-center components in a framework very similar to that of calibration model 2, where both levels of effect were estimated simultaneously. In this type of disease model, the between-center component has the great advantage over a “classical” ecologic study that it can adjust for confounding at the individual level and over an aggregate design (35) and that the populations under study provide estimates of cancer incidence and dietary exposure.

The impact of energy adjustment on observed macronutrient variability appeared to be considerable in the calibration model. Adjustment for energy is motivated both by the requirement to consider isocaloric models and by the partial control of measurement error that it can bring about (36). Within-center variances were affected more than between-center variance, particularly for carbohydrates and lipids, reflecting a reduction in the error variance. Thus, the ICC values for the macronutrient densities were considerably larger than those for the absolute intakes.

The calibration model relies on some essential statistical assumptions. First, the errors in R and Q measurements must be uncorrelated. In practice, this assumption may not be fully met, and some positive correlation between errors may occur, for instance, because of subjects’ tendencies to under- or overreport intake when different dietary assessment methods are used (14, 16, 37, 38). A positive correlation between errors leads to overestimation of the within-center calibration coefficient λB and thus to an underestimation of shrinkage of the intake distributions within centers. The ICC values in the present analyses may thus underestimate the relative weight that the between-center component should have in a full multilevel analysis. There is some evidence that adjustment for total energy intake, either through a linear model or by calculating nutrient densities, can reduce the magnitude of correlation between random measurement errors of Q measurements and R or food consumption records (36, 39). The effective error variance for the macronutrients is also probably reduced by energy adjustment. This seems to happen because the error in energy and the error in estimated macronutrient intake are highly correlated and, after adjustment, partly “cancel” each other (18). Thus, although the ICC values for the absolute intakes may be considerably underestimated, the ICC values for the energy-adjusted macronutrients can be less affected. The latter are of greater relevance, since some form of energy adjustment of macronutrient intakes is standard practice in epidemiology.

A second major assumption is that R provide unbiased estimates of dietary intake at a group level; that is, αp = 0 and βp = 1.0 in measurement-error model 1b (absence of constant and proportional scaling biases). This assumption may be relaxed, however, to that of homogeneous scaling biases (i.e., homogeneous αp’s and βp’s) in all study centers, which in practice may be reached approximately. Again, energy adjustment may help to standardize dietary measurements and meet at least the relaxed assumption. Under this assumption, the measurement error correction will still provide a correct relative weighting of within- versus between-center evidence for diet-disease associations. In addition, βp’s of less than unity lead to underestimation of within-center calibration factors. In this context, it is possible that the correlation between errors and the reference measurement bias may counterbalance their effects, thus leading to correct estimates of λB.

The calibration model in model 2 assumes fully linear associations between response and explanatory variables. Nonlinear calibration models are advisable solutions to capture the lack of linearity between R and Q. The shrinkage of the within-center variance component would characterize the predicted macronutrient distribution, also in nonlinear calibration models.

Measurement errors bias the association between any two quantities, for example, the relative risk estimate in a disease model, and cause power loss. Within single study centers, calibration may correct for bias but does not recover power losses due to random errors. Indeed, the effect of shrinkage of exposure distributions generally outweighs the effect of bias correction (i.e., increase) in relative risk estimates. In a multicenter study, however, calibration in principle can reduce between-center heterogeneity in diet-disease associations caused by differential impact of measurement errors and, probably more importantly, will lead to a less biased estimate of the overall relative risk (7, 11).

Moreover, in estimating associated variability of corrected relative risk estimates, it is necessary to take into account the variability in the estimate λB’s (10, 11). In the EPIC study, this issue is minor because the large size of the calibration study allows the calibration factor to be estimated with high precision.

ACKNOWLEDGMENTS

The EPIC study was supported by the “Europe Against Cancer” Programme of the European Commission (SANCO); Ligue contre le Cancer (France); Société 3M (France); Mutuelle Générale de l’Education Nationale; Institut National de la Santé et de la Recherche Médicale (INSERM); German Cancer Aid; German Cancer Research Center; German Federal Ministry of Education and Research; Danish Cancer Society; Health Research Fund (FIS) of the Spanish Ministry of Health; the participating regional governments and institutions of Spain; Cancer Research UK; Medical Research Council, United Kingdom; the Stroke Association, United Kingdom; British Heart Foundation; Department of Health, United Kingdom; Food Standards Agency, United Kingdom; the Wellcome Trust.

United Kingdom; Greek Ministry of Health; Greek Ministry of Education; Italian Association for Research on Cancer; Italian National Research Council; Dutch Ministry of Public Health, Welfare and Sports; Dutch Ministry of Health; Dutch Prevention Funds; LK Research Funds; Dutch ZON (Zorg Onderzoek Nederland); World Cancer Research Fund (WCRF); Swedish Cancer Society; Swedish Scientific Council; Regional Government of Skane, Sweden; and the Norwegian Cancer Society.

REFERENCES


