Implications of a New Dietary Measurement Error Model for Estimation of Relative Risk: Application to Four Calibration Studies

Victor Kipnis,1 Raymond J. Carroll,2 Laurence S. Freedman,3 and Li Li4

Food records or 24-hour recalls are currently used to calibrate food frequency questionnaires (FFQs) and to correct disease risks for measurement error. The standard regression calibration approach requires that these reference measures contain only random within-person errors uncorrelated with errors in FFQs. Increasing evidence suggests that records/recalls are likely to be also flawed with systematic person-specific biases, so that for any individual the average of multiple replicate assessments may not converge to her/his true usual nutrient intake. The authors propose a new measurement error model to accommodate person-specific bias in the reference measure and its correlation with systematic error in the FFQ. Sensitivity analysis using calibration data from four studies demonstrates that failure to account for person-specific bias in the reference measure can often lead to substantial underestimation of the relative risk for a nutrient. These results indicate that in the absence of information on the extent of person-specific biases in reference instruments and their relation to biases in FFQs, the adequacy of the standard methods of correcting relative risks for measurement error is in question, as is the interpretation of negative findings from nutritional epidemiology such as failure to detect an important relation between fat intake and breast cancer. Am J Epidemiol 1999; 150:642-51.

dietary assessment methods; epidemiologic methods; measurement error; models, statistical; nutrient intake; regression analysis

In nutritional epidemiologic studies, long-term habitual dietary intake is usually the primary exposure of interest. Researchers have long recognized that dietary measurement is subject to substantial error that can have a profound impact on assessment of the effect of an exposure on disease. The interpretation of findings from nutritional epidemiology is therefore critically dependent on the assessment of, and adjustment for, dietary measurement error.

This can be exemplified by the controversy over the possible association between dietary fat and breast cancer. Although findings from animal research, international correlational, and case-control studies demonstrate a strong positive relation between fat intake and breast cancer, cohort studies do not. A pooled analysis of seven cohort studies (1) reported a statistically nonsignificant energy-adjusted relative risk of only 1.02 for a change of 25 g/day in fat intake. This inconsistency may be attributable to a lack of association between dietary fat and breast cancer, the international correlational and case-control studies being explained by unmeasured confounding and recall bias, respectively. However, it may also be due to methodological limitations, in particular the assessment of dietary error and appropriate adjustment for it.

The cited reports are based on data from food frequency questionnaires (FFQs) which measure a person's usual dietary intake over a defined period. Because the FFQ is relatively inexpensive and easy to administer, it has become the tool of choice for large-scale nutritional epidemiologic studies. Many investigators have recognized that data captured by FFQs are subject to considerable error. Usually, measurement error in an exposure variable attenuates (biases toward zero) the estimates of disease risk for that exposure.
and reduces statistical power for the corresponding significance test. An important direct relation between diet and disease, therefore, may be obscured.

Researchers have attempted to estimate the attenuation factor to apply a correction and get unbiased estimates of disease risk. Initially, it was assumed that measurement errors were associated with within-person random, rather than systematic, variation (2–5). In nutritional studies, it was later realized that measurement errors in FFQs may contain systematic bias and be correlated with true exposure (6, 7).

To obtain a corrected risk estimate, Rosner et al. (6) introduced the linear regression calibration approach, which estimates the attenuation coefficient as the slope of true on observed exposure. The method requires subsample reference measurements which either represent true exposure or, if they are themselves imperfect and contain error, have error that is completely random and independent of the error in the FFQ. Because this approach allows for systematic bias as well as within-person random variation in the FFQ, it has gained recognition as the best currently available approach for correcting risk estimates for dietary measurement error. Its application has promoted the integration of calibration substudies in large epidemiologic investigations that involve more expensive and time consuming dietary assessment methods, such as a series of multiple-day food records or 24-hour recalls, as reference measurements.

In the pooled cohort studies (1), correction for measurement error using this method did not materially alter the findings: the pooled corrected relative risk for energy-adjusted total fat increased from its “uncorrected” estimate of 1.02 to 1.07 (per 25 g; 95 percent confidence interval 0.86–1.34). As indicated above, however, the applied measurement model imposes certain important requirements on the reference measurements. Prentice (8), using the Women’s Health Trial Vanguard Study as an external calibration study and adopting a different measurement error model, concluded that the pooled cohort data could be consistent with the findings from the international correlation analysis.

In the present paper, we investigate the robustness of the measurement model adopted in the standard regression calibration approach to some likely violations of modeling assumptions. Based on recent evidence, we first develop a new measurement error model that allows for systematic person-specific bias in a reference instrument that may be correlated with its counterpart in a FFQ. We then apply the new and standard models to data from four calibration studies and compare the resultant corrections to risk estimates.

MODELS AND METHODS
Regression calibration approximation

Consider the disease model

\[ R(D|T) = \alpha_0 + \alpha_1 T, \quad (1) \]

where \( R(D|T) \) is the risk of disease \( D \) on an appropriate scale (e.g., logit or logarithmic), \( T \) denotes true long-term usual intake of a given nutrient, \( \alpha_0 \) is the intercept, and \( \alpha_1 \) is the slope and represents a measure of the disease risk associated with intake of the nutrient. Instead of \( T \), we observe \( Q \), the nutrient intake obtained from a FFQ. For a given individual, measurement error \( e_Q \) is defined as the difference between \( Q \) and \( T \), i.e., \( Q = T + e_Q \). Note that short-term variation in diet is included in \( e_Q \), as well as random and systematic error components resulting from the instrument itself. We assume throughout that error \( e_Q \) is nondifferential with respect to disease \( D \); i.e., \( Q \) contributes no additional information about \( D \) beyond what is available in \( T \).

Regressing \( D \) on observed intake \( Q \), instead of true intake \( T \), produces a biased estimate of the disease risk (9, 10). One way of adjusting this estimate for measurement error, known as regression calibration, is to substitute for the unknown true intake its best predicted value given the measured intake, or \( E(T|Q) \) (6, 11–18). If the regression calibration model is linear, i.e.,

\[ E(T|Q) = \lambda_0 + \lambda_1 Q, \quad (2) \]

then, to an excellent approximation,

\[ R(D|Q) = \gamma_0 + \gamma_1 Q, \quad (3) \]

where \( \gamma_1 = \lambda_1 \alpha_1 \).

Thus, the “naive” risk coefficient \( \gamma_1 \) in model 3 needs to be divided by \( \lambda_1 \) to recover the true coefficient \( \alpha_1 \) in model 1. Factor \( \lambda_1 \) is the slope of the linear regression calibration model 2, and its value is given by

\[ \lambda_1 = \text{cov}(T,Q)/\text{var}(Q). \quad (4) \]

Although, in principle, when measurement error \( e_Q \) is correlated with true exposure \( T \), \( \lambda_1 \) could be negative or greater than one in magnitude, in nutritional studies usually \( \lambda_1 \) lies between 0 and 1 (19) and can be thought of as an attenuation of the coefficient \( \alpha_1 \).
Adjustment for measurement error

The attenuation factor $\lambda_1$ is usually estimated from a smaller calibration substudy. In the ideal case, the true exposure variable $T$ can be precisely measured in the calibration study and regressed on its surrogate $Q$ to obtain the estimate $\hat{\lambda}_1$. Unfortunately, in dietary studies, the perfect "gold standard" for true nutrient intake does not exist. The standard regression calibration approach (6), however, requires only that there be a reference instrument $F$ such that the slope of the linear regression of $F$ on $Q$ be equal to $\lambda_1$ or, as follows from formula 4, that

$$\text{cov}(F,Q) = \text{cov}(T,Q)$$

Writing the measurement model as

$$Q = T + e_Q$$
$$F = T + e_F$$

assumption 5 is satisfied if

$$\text{cov}(e_F,T) = 0,$$
$$\text{cov}(e_F,e_Q) = 0.$$  \hspace{1cm} (8, 9)

The attenuation factor $\lambda_1$ is then estimated as

$$\hat{\lambda}_1 = \frac{\text{cov}(F,Q)}{\text{var}(Q)}.$$  \hspace{1cm} (5)

Allowing for any error component such as systematic bias or within-person random variation in a FFQ, measurement model 6–9, nevertheless, adopts relatively strong assumptions 8–9 regarding the reference measurements in the calibration study. Although a reference instrument may contain errors, these errors should be uncorrelated with both true intake and any error components in the FFQ. Based on recent evidence, we question these assumptions and will investigate the robustness of conclusions to their violation. Multiple-day food records and/or 24-hour recalls have been traditionally used as a reference instrument. Below, we develop a model that seems more appropriate for such instruments.

A new dietary measurement error model

**Model for the FFQ.** The error $e_Q$ in a FFQ is thought likely to include a systematic bias $b$ that may depend on the individuals' true intake $T$, as well as within-person variation $\varepsilon$ (7, 19–20). We model the relation between bias $b$ and true intake $T$ as the linear regression

$$b = \beta_0 + \beta_1 T + r,$$

where $r$ has zero mean and is independent of $T$. The component

$$\beta_0 + \beta_1 T$$

is common to all persons with the same true intake and may be called group-specific bias. It can be thought of as arising from correlation between error and true intake. For example, given the social/cultural pressure to follow the "correct" dietary pattern, persons with a low intake of supposedly healthy food may be tempted to overreport intake, and those with a high intake of supposedly unhealthy food may be tempted to underreport.

The difference $r$ between within-person bias and its group-specific component may be caused by personality characteristics such as susceptibility to social/cultural influences, and will be called person-specific bias. Note that it is part of within-person systematic error and will be reproduced in repeated measurements on the same individual.

Gathering all the error components together, we write the reported intake $Q_{ij}$ of the $i$th among a group of $n$ individuals at the $j$th of $m_Q$ repeated measurements, as

$$Q_{ij} = \beta_0 + \beta_1 T_i + r_i + \varepsilon_{ij},$$

where

$$\beta_1 = \beta_1^* + 1.$$  \hspace{1cm} (Am J Epidemiol Vol. 150, No. 6, 1999)

**Model for the reference instrument.** In the standard application of the regression calibration approach, it is assumed that there is no bias in the reference instrument, i.e., if we took enough repeated 24-hour recalls or multiple-day food records, we could get as close as we like to the true long-term intake for each person. Recent evidence suggests that a reappraisal of this assumption is now needed. Studies of doubly-labeled water and urinary nitrogen (21–27) indicate that reports using recalls or food records are biased (on average toward under-reporting), and that individuals systematically differ in their reporting accuracy. Thus, we need to allow for systematic person-specific bias in the reference instruments as well as in the FFQs. The simplest such model for a reference instrument can be written as

$$F_{ij} = \mu_0 + T_i + s_i + u_{ij}, \quad j = 1, \ldots, m_F.$$
where \( \mu_0 \) represents group-specific bias, and \( s \) and \( u \) denote person-specific bias and within-person random error, respectively, and are assumed to be independent of each other and of true intake \( T \), and \( m_r \) denotes the number of repeat measurements.

**The full model.** Combining expressions for error structure in a FFQ and a reference instrument, the new measurement model is written as

\[
Q_{ij} = \beta_0 + \beta_i T_i + r_i + \varepsilon_{ij}, \quad j = 1, \ldots, m_Q \geq 2, \tag{10}
\]

\[
F_{ij} = \mu_0 + T_i + s_i + u_{ij}, \quad j = 1, \ldots, m_F \geq 2, \tag{11}
\]

where \( T \) has mean \( \mu_T \) and variance \( \sigma_T^2 \) and \( r, s, \varepsilon, \) and \( u \) have mean zero and variances \( \sigma_r^2, \sigma_s^2, \sigma_\varepsilon^2, \) and \( \sigma_u^2 \), respectively. All random variables on the right-hand side of equations 10–11 are assumed to be mutually independent with two important exceptions. First, fluctuation of diet over time may induce correlation between random within-person errors of different measurements taken at the same time. Thus, we allow for

\[
\text{corr}(\varepsilon_{ij}, u_{ij}) \neq 0,
\]

if the main and reference instruments are administered contemporaneously. Second, person-specific biases, \( r_i \) and \( s_i \), are likely to share a source common to both self-reporting instruments, such as personal susceptibility to social pressure to have a healthy diet. This will induce positive correlation between them and, therefore, we allow

\[
\rho_{r,s} = \text{corr}(r, s) > 0.
\]

Comparing the new model with the standard one, note that total error \( e_F = \mu_0 + s + u \) in the reference instrument \( F \) is still assumed to be uncorrelated with true intake \( T \). Thus, the first of the two main requirements involved in the standard approach, assumption 8, holds under the new model. However, correlations between within-person random errors \( \varepsilon \) and \( u \) and, especially, between person-specific biases \( r \) and \( s \) violate assumption 9. Therefore, fitting the new model to the same data will allow us to compare the results with that of the standard approach, and to assess the robustness of the standard approach to such correlations.

Unfortunately, not all the parameters in model 10–11 can be identified without additional information because there are 11 parameters, but only eight uniquely identified first and second moments. To compare the results with the standard approach, we can, however, perform a sensitivity analysis by adopting a series of prespecified values for three of the parameters and exploring sensitivity of the results to changes in these values.

**Comparison with other models**

Recently, there have been suggested several alternatives to the measurement model 6–9 adopted in the standard regression calibration approach. Freedman et al. (7) proposed a model that allows for correlated within-person random errors in a FFQ and a reference instrument administered close in time, but assumes no bias in \( F \). Plummer and Clayton (20) and Spiegelman et al. (28) also allow for correlated within-person random errors in \( Q \) and \( F \), even when the two measures are taken apart, but do not consider person-specific biases. In particular, Spiegelman et al. (28) demonstrate using several examples that, although their modified model leads to a different estimated risk coefficient compared with the standard approach, the difference is small and materially unimportant. As shown in Appendix 1, ignoring person-specific bias in a reference instrument always leads to a greater estimated attenuation factor compared with our model, and therefore accounts for only part of the attenuation if, in fact, this bias takes place.

Kaaks et al. (19) were the first to consider person-specific biases in both a FFQ and a reference instrument, but assumed that these biases were uncorrelated. Prentice (8) suggested a model that is very close to ours in spirit. It allows for group-specific bias in a FFQ related to true intake and correlated person-specific biases in both a FFQ and a reference instrument. In addition, the model parameters are allowed to depend on body mass index tertiles, making the model somewhat more general. However, as shown in Appendix 2, this model involves an implicit assumption that \( \rho_{r,s} = \sigma_r/\sigma_s \). This condition imposes strict boundaries on the joint distribution of person-specific biases in a FFQ and a reference instrument that may not be warranted by data.

**SENSITIVITY ANALYSIS USING FOUR CALIBRATION STUDIES**

**Data**

Calibration data on fat nutrient density (percent of energy from fat) from four studies were used: the Women's Health Trial Vanguard Study (WHT), the Nurses' Health Study (NHS), the National Institutes of Health - American Association of Retired Persons Diet and Health Study (AARP), and the American Cancer Society Prevention II (CPS II) Nutrition Survey Validation and Red Meat Consumption Study (ACS) (29–32).
In the WHT study, women volunteers, aged 45–69 years, were randomized into the intervention and control groups. We used data on 86 subjects in the control group, which consisted of three 4-day food records (at 6 months, 1 year, and 2 years after entry into the study) and two semi-quantitative FFQs (at 1 year and 2 years) developed by Block et al. (33).

In the NHS study, a 61-item Willett’s FFQ was administered at baseline in 1980. Healthy nurses who resided in the Boston area, aged 34–59 years, were selected from among the responders to the baseline FFQ into the validation study and completed another FFQ one year later. In addition, each participant completed four 7-day food records at approximately 3-month intervals beginning 3 months after administration of the baseline FFQ. We analyzed complete records on 168 women.

In the AARP study, an extensive FFQ was mailed in 1995–1996 to 3.5 million AARP members, aged 50–69 years, in six states (California, Florida, Louisiana, New Jersey, North Carolina, and Pennsylvania) and two cities (Atlanta, Georgia, and Detroit, Michigan). The baseline FFQ data were used to classify participants of each gender into five strata for each of the four dietary variables: percent of energy from fat, dietary fiber, servings of fruit and vegetables, and servings of red meat. In order to increase the subsample variability of these four variables, eligible participants were chosen into the calibration study by over-sampling extreme values of these four variables, with fewer individuals sampled from strata 2 and 4 and still fewer from stratum 3. The participants in the calibration study were then administered two 24-hour recalls over the telephone (on average one month apart) and, several months later, were mailed the second FFQ. We analyzed data on 969 women in the calibration study.

In the ACS study, participants, men and women, aged 50–69 years, were recruited for the validation study. Eligibility requirements included no major dietary changes or weight loss within the previous year based on the CPS II Nutrition Survey FFQ. Four 24-hour recalls were administered in person at approximately 3-month intervals over the period of one year and resulted in an approximately three-to-one ratio of week to weekend days. A repeat FFQ was then self-administered and completed. Complete data on 184 women were used for the present analysis.

Model fitting

As mentioned above, to fit the model and perform the sensitivity analysis we need to prespecify values for three out of 11 parameters. Based on indications in the literature (24, 34) that nutrient density may not contain group-specific bias, we put \( \mu_0 = 0 \). Note that this assumption only affects the estimate of mean true intake, \( \mu_t \), and does not change the estimated risk coefficient. The main distinction of the new model compared with the standard one is the presence of a person-specific bias in the reference instrument which is allowed to be correlated with its counterpart in the FFQ. We, therefore, study the sensitivity of the results to the ratio \( k = \sigma_{ef}^2 / \sigma_{rt}^2 \) of the variances of person-specific biases and the correlation coefficient \( \rho_{rs} \) between them.

Because a reference instrument is likely to be no less precise than a FFQ, we considered \( k = 0.25; 0.5; 1.0 \). To allow for possibly strong correlation, we varied \( \rho_{rs} \) from 0.0 to 0.6 with steps of 0.1. For comparison purposes, we also considered the case of \( k = 0 \) (i.e., \( \sigma_{ef}^2 = 0 \)), which corresponds to the measurement error model considered by Freedman et al. (7). The remaining parameters of the model were estimated using the maximum likelihood method under normal distribution, a reasonable assumption for fat nutrient density. We also checked the validity of the estimates by the bootstrap method with 500 replicates that produced essentially the same results.

To further illustrate the effect of measurement error on the estimated relative risk (RR), we also estimated the true (deattenuated) RR assuming that the observed RR using the error-prone FFQ was 1.1, i.e., a value similar to that reported by Hunter et al. (1). This estimate, denoted by \( \text{RR}_{TIO} \), was calculated as

\[
\text{RR}_{TIO} = 1.1^k.
\]

Results

The descriptive statistics for reported percent of energy from fat from the considered studies are presented in table 1. Of note are the mean values in the AARP and ACS studies, which are noticeably smaller than in the other two studies. The standard deviations in the NHS are noticeably smaller for both the FFQ and reference instrument, which reflects the relative homogeneity of reported intake in this population.

Table 2 provides the estimated attenuation factor \( \lambda \) for different values of \( k = \sigma_{ef}^2 / \sigma_{rt}^2 \) and \( \rho_{rs} \). We emphasize that these estimates are from a sensitivity analysis. We can show that if \( \rho_{rs} = 0 \), the attenuation factor does not depend on the ratio \( k = \sigma_{ef}^2 / \sigma_{rt}^2 \), so its estimate is provided in the first row for \( k = 0 \). From the table, the pattern of change in \( \lambda \) seems to depend very little on \( k \). The attenuation factor decreases (attenuation increases) rather slowly when \( \rho_{rs} \) increases from 0 to 0.3, but its value drops steeply when \( \rho_{rs} \) continues to increase to 0.6 in the WHT, NHS, and

Am J Epidemiol Vol. 150, No. 6, 1999
TABLE 1. Descriptive statistics for percent of energy from fat measured in the four calibration studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Dietary assessment Instrument</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHT* (n = 86)</td>
<td>FFQ* at 1 year</td>
<td>37.13</td>
<td>8.94</td>
</tr>
<tr>
<td></td>
<td>FFQ at 2 years</td>
<td>36.37</td>
<td>8.31</td>
</tr>
<tr>
<td></td>
<td>4-day food record at 6 months</td>
<td>37.12</td>
<td>7.77</td>
</tr>
<tr>
<td></td>
<td>4-day food record at 1 year</td>
<td>36.42</td>
<td>8.05</td>
</tr>
<tr>
<td></td>
<td>4-day food record at 2 years</td>
<td>38.26</td>
<td>7.99</td>
</tr>
<tr>
<td>NHS* (n = 168)</td>
<td>FFQ at baseline</td>
<td>37.90</td>
<td>7.39</td>
</tr>
<tr>
<td></td>
<td>FFQ repeated</td>
<td>36.92</td>
<td>6.73</td>
</tr>
<tr>
<td></td>
<td>7-day food record (week 1)</td>
<td>37.77</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>7-day food record (week 2)</td>
<td>37.72</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td>7-day food record (week 3)</td>
<td>38.37</td>
<td>5.35</td>
</tr>
<tr>
<td></td>
<td>7-day food record (week 4)</td>
<td>37.92</td>
<td>5.29</td>
</tr>
<tr>
<td>AARP* (n = 969)</td>
<td>FFQ at baseline</td>
<td>29.95</td>
<td>8.45</td>
</tr>
<tr>
<td></td>
<td>FFQ repeated</td>
<td>30.51</td>
<td>8.04</td>
</tr>
<tr>
<td></td>
<td>24-hour recall 1</td>
<td>30.27</td>
<td>10.25</td>
</tr>
<tr>
<td></td>
<td>24-hour recall 2</td>
<td>30.68</td>
<td>10.02</td>
</tr>
<tr>
<td>ACS* (n = 184)</td>
<td>FFQ at baseline</td>
<td>33.78</td>
<td>9.42</td>
</tr>
<tr>
<td></td>
<td>FFQ repeated</td>
<td>32.81</td>
<td>9.03</td>
</tr>
<tr>
<td></td>
<td>24-hour recall at 3 months</td>
<td>29.67</td>
<td>9.86</td>
</tr>
<tr>
<td></td>
<td>24-hour recall at 6 months</td>
<td>30.42</td>
<td>9.15</td>
</tr>
<tr>
<td></td>
<td>24-hour recall at 9 months</td>
<td>30.04</td>
<td>9.27</td>
</tr>
<tr>
<td></td>
<td>24-hour recall at 12 months</td>
<td>30.08</td>
<td>8.66</td>
</tr>
</tbody>
</table>

* WHT, Women's Health Trial Vanguard Study; NHS, Nurses’ Health Study; AARP, National Institutes of Health - American Association of Retired Persons Diet and Health Study; ACS, American Cancer Society Prevention II (CPS II) Nutrition Survey Validation and Red Meat Consumption Study; FFQ, food frequency questionnaire.

AARP studies. In the ACS study, the change in \( \lambda \), although relatively more accelerated for higher values of \( \rho_{r,s} \), remains fairly modest. The sensitivity of \( \lambda \) to \( \rho_{r,s} \) is somewhat less strong when \( k = 0.25 \) for the WHT, NHS, and AARP studies compared with \( k = 0.5 \) and \( k = 1.0 \).

Table 3 provides the comparison of the estimated attenuation factors and the corresponding deattenuated RRs assuming an observed RR of 1.1 between the standard model and the new model with \( \rho_{r,s} \) equal to 0.0, 0.3, and 0.6 and \( k = 1.0 \). It follows that the presence of person-specific bias in the reference instrument that is uncorrelated with person-specific bias in the FFQ (\( \rho_{r,s} = 0 \)) does not change the results very much. The difference between the two models becomes more important if \( \rho_{r,s} = 0.3 \), and the results from the new model are strikingly different if \( \rho_{r,s} \) reaches the value of 0.6. At this level of \( \rho_{r,s} \) for the WHT and NHS studies, the deattenuated RR is well in line with the values from the international correlational studies of fat and breast cancer, and for the AARP study it is somewhat lower but still substantial. In contrast, for the ACS study, the deattenuated RR is essentially unchanged by the new model. It is worth noting, however, that when we increased the value of \( \rho_{r,s} \) to 0.8, the corrected RR in the ACS study was equal to 1.98.

From table 3, one can see that the estimated attenuation factor and the corresponding corrected RR vary among studies. It is of interest, therefore, to compare the relative change in the estimated attenuation factor when \( \rho_{r,s} \) increases from 0.0 to 0.3 to 0.6. Table 4 demonstrates that the increase in \( \rho_{r,s} \) leads to the same relative decrease in the attenuation factor in the NHS and AARP studies and a slightly greater, although not statistically significantly different, change in the WHT study. The ACS study remains different, displaying statistically significantly smaller relative reductions in the estimated attenuation factor than the other studies.

**DISCUSSION**

Recently, there has been a growing awareness that person-specific bias in self-reported intake represents a major challenge now facing nutritional studies. Such bias can affect the analysis and interpretation of all nutritional studies relying on self-report. In the area of correcting disease RR estimates that we have considered in this paper, the correlation between person-specific biases in a reference instrument, such as food records/recalls, and a FFQ violates one of the main assumptions of the standard regression calibration approach. In this paper, we consider a new model that, for a reference instrument, incorporates person-specific bias and allows
TABLE 2. Estimated attenuation factor ($\lambda$) for percent energy from fat in four calibration studies as a function of the ratio of the variances of person-specific biases ($k = \sigma_i^2 / \sigma_i^*$) in the reference measurement and FFQ* and the correlation $\rho_{ir}$ between these biases

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\rho_{ir}$</th>
<th>WHT*</th>
<th>NHS*</th>
<th>AARP*</th>
<th>ACS*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.1</td>
<td>0.34 (0.06)</td>
<td>0.26 (0.04)</td>
<td>0.47 (0.03)</td>
<td>0.49 (0.03)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.32 (0.06)</td>
<td>0.25 (0.04)</td>
<td>0.45 (0.03)</td>
<td>0.38 (0.04)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.30 (0.06)</td>
<td>0.23 (0.04)</td>
<td>0.43 (0.03)</td>
<td>0.37 (0.04)</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>0.27 (0.06)</td>
<td>0.21 (0.05)</td>
<td>0.40 (0.03)</td>
<td>0.36 (0.05)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.23 (0.06)</td>
<td>0.19 (0.05)</td>
<td>0.37 (0.04)</td>
<td>0.34 (0.05)</td>
</tr>
<tr>
<td>0.6</td>
<td>0.19 (0.06)</td>
<td>0.16 (0.05)</td>
<td>0.33 (0.04)</td>
<td>0.31 (0.05)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.34 (0.06)</td>
<td>0.26 (0.04)</td>
<td>0.46 (0.03)</td>
<td>0.39 (0.04)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.32 (0.06)</td>
<td>0.25 (0.04)</td>
<td>0.44 (0.03)</td>
<td>0.38 (0.04)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.28 (0.06)</td>
<td>0.23 (0.04)</td>
<td>0.41 (0.03)</td>
<td>0.37 (0.04)</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>0.24 (0.06)</td>
<td>0.20 (0.05)</td>
<td>0.38 (0.04)</td>
<td>0.35 (0.05)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.19 (0.06)</td>
<td>0.16 (0.06)</td>
<td>0.33 (0.04)</td>
<td>0.33 (0.05)</td>
</tr>
<tr>
<td>0.6</td>
<td>0.11 (0.07)</td>
<td>0.11 (0.07)</td>
<td>0.26 (0.05)</td>
<td>0.30 (0.06)</td>
<td></td>
</tr>
</tbody>
</table>

WHT*, Women’s Health Trial Vanguard Study; NHS*, Nurses’ Health Study; AARP*, National Institutes of Health - American Association of Retired Persons Diet and Health Study; ACS*, American Cancer Society Prevention II (CPS II) Nutrition Survey Validation and Red Meat Consumption Study. 

* FFQ, food frequency questionnaire; WHT, Women’s Health Trial Vanguard Study; NHS, Nurses’ Health Study; AARP, National Institutes of Health - American Association of Retired Persons Diet and Health Study; ACS, American Cancer Society Prevention II (CPS II) Nutrition Survey Validation and Red Meat Consumption Study.

\[ k = \frac{\sigma_i^2}{\sigma_i^*} \]

\[ \rho_{ir} \]

\[ * \] Standard deviation in parentheses.

TABLE 3. Estimated attenuation factor ($\lambda$) and deattenuated relative risk (RR\textsubscript{TO}) assuming that the observed RR using a FFQ* is equal to 1.10 for the standard model and the new model as a function of the correlation ($\rho_{ir}$) between person-specific biases in the FFQ and reference measurement for percent of energy from fat in four calibration studies (for $\sigma_i^2 / \sigma_i^* = 1$)

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>WHT*</th>
<th>NHS*</th>
<th>AARP*</th>
<th>ACS*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>$\lambda$</td>
<td>0.42 (0.32, 0.52)</td>
<td>0.31 (0.23, 0.39)</td>
<td>0.50 (0.45, 0.55)</td>
<td>0.60 (0.51, 0.90)</td>
</tr>
<tr>
<td></td>
<td>$\text{RR}_{\text{TO}}$</td>
<td>1.25 (1.20, 1.35)</td>
<td>1.36 (1.28, 1.51)</td>
<td>1.21 (1.19, 1.24)</td>
<td>1.27 (1.22, 1.36)</td>
</tr>
<tr>
<td>New $\rho_{ir} = 0.0$</td>
<td>$\lambda$</td>
<td>0.36 (0.26, 0.48)</td>
<td>0.27 (0.21, 0.32)</td>
<td>0.48 (0.43, 0.53)</td>
<td>0.40 (0.33, 0.47)</td>
</tr>
<tr>
<td></td>
<td>$\text{RR}_{\text{TO}}$</td>
<td>1.30 (1.22, 1.44)</td>
<td>1.42 (1.35, 1.57)</td>
<td>1.22 (1.20, 1.29)</td>
<td>1.27 (1.21, 1.33)</td>
</tr>
<tr>
<td>$\rho_{ir} = 0.3$</td>
<td>$\lambda$</td>
<td>0.29 (0.21, 0.42)</td>
<td>0.23 (0.16, 0.30)</td>
<td>0.41 (0.35, 0.48)</td>
<td>0.37 (0.29, 0.46)</td>
</tr>
<tr>
<td></td>
<td>$\text{RR}_{\text{TO}}$</td>
<td>1.39 (1.25, 1.57)</td>
<td>1.51 (1.37, 1.81)</td>
<td>1.26 (1.22, 1.31)</td>
<td>1.29 (1.23, 1.39)</td>
</tr>
<tr>
<td>$\rho_{ir} = 0.6$</td>
<td>$\lambda$</td>
<td>0.14 (0.06, 0.29)</td>
<td>0.16 (0.06, 0.25)</td>
<td>0.24 (0.14, 0.35)</td>
<td>0.22 (0.1, 0.42)</td>
</tr>
<tr>
<td></td>
<td>$\text{RR}_{\text{TO}}$</td>
<td>1.96 (1.39, 4.90)</td>
<td>1.96 (1.46, 4.90)</td>
<td>1.48 (1.31, 1.97)</td>
<td>1.34 (1.25, 1.57)</td>
</tr>
</tbody>
</table>

* FFQ, food frequency questionnaire; WHT, Women’s Health Trial Vanguard Study; NHS, Nurses’ Health Study; AARP, National Institutes of Health - American Association of Retired Persons Diet and Health Study; ACS, American Cancer Society Prevention II (CPS II) Nutrition Survey Validation and Red Meat Consumption Study.

\[ * \] 95% confidence interval in parentheses.

this bias to be correlated with its counterpart in a FFQ. Unfortunately, the model is not identifiable without some additional information and, at the moment, can only be used in a sensitivity analysis. We performed such an analysis by considering prespecified values for the ratio $k = \sigma_i^2 / \sigma_i^*$ of the variances of person-specific biases in a reference instrument and FFQ and the correlation $\rho_{ir}$ between these biases. We applied the new model, as well as the standard one, to four calibration data sets to check whether the measurement error adjustment to the estimate of relative risk from fat depends on the measurement error model used.

Our study suggests that a large correlation $\rho_{ir}$ between the person-specific biases in a FFQ and a ref-

Am J Epidemiol Vol. 150, No. 6, 1999
TABLE 4. Bootstrap estimated percent reduction in the attenuation factor due to increase in the correlation (ρ_u) between person-specific biases in the FFQ and reference measurement for σ^2/σ_u^2 = 1 for percent energy from fat in four calibration studies

<table>
<thead>
<tr>
<th>Increase in ρ_u</th>
<th>WHT* Mean (SD*)</th>
<th>NHS* Mean (SD)</th>
<th>AARP* Mean (SD)</th>
<th>ACS* Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 to 0.3</td>
<td>19 (5)</td>
<td>16 (5)</td>
<td>15 (3)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>0.3 to 0.6</td>
<td>47 (16)</td>
<td>39 (14)</td>
<td>39 (10)</td>
<td>12 (7)</td>
</tr>
<tr>
<td>0.0 to 0.6</td>
<td>56 (15)</td>
<td>48 (15)</td>
<td>48 (10)</td>
<td>17 (9)</td>
</tr>
</tbody>
</table>

* FFQ, food frequency questionnaire; WHT, Women's Health Trial Vanguard Study; NHS, Nurses' Health Study; AARP, National Institutes of Health - American Association of Retired Persons Diet and Health Study; ACS, American Cancer Society Prevention II (CPS II) Nutrition Survey Validation and Red Meat Consumption Study; SD, standard deviation.

dereference instrument may lead to strikingly different results. If ρ_{rs} is smaller than 0.3, the standard regression calibration model 6–9 appears to be rather robust to the presence of person-specific bias in a reference instrument, accounting for most of the attenuation of RR due to measurement error. But for moderate/strong correlation (ρ_{rs} ≥ 0.3), the situation changes dramatically for the three out of four data sets considered. Indeed, in the WHT, NHS, and AARP studies, the standard approach adjusts for 81 percent to 85 percent of attenuation of RR due to measurement error when ρ_{rs} = 0.3 and only for 44 percent to 52 percent when ρ_{rs} reaches 0.6. As demonstrated in table 4, this may lead to serious underestimation of large and potentially biologically important effects of dietary intake on disease. However, when we applied the new model to the ACS data, major differences with the standard model arose only when ρ_{rs} became 0.8. Thus, whether accounting for correlation between person-specific biases in a reference instrument and a FFQ substantially changes the adjusted relative risk appears to depend not only on the magnitude of this correlation, but potentially on the particular study used to perform the assessment. If six other validation studies used in the pooled analysis by Hunter et al. (1) are similar to the seventh study, the NHS, the observed relative risk of 1.1 between the highest and lowest quintiles of fat intake may be compatible with the results of the international correlational studies. On the other hand, if the majority of these studies mirrors the ACS data, the deattenuated relative risk estimated with the standard model may truly reflect the absence of any fat effect. Thus, our results suggest that the calibration studies used by Hunter et al. (1) should, where possible, be subjected to a sensitivity analysis of the type demonstrated here.

Clearly, it is important to know the correlation ρ_{rs} between person-specific biases in a FFQ and reference instrument. At this stage, there are no data known to us to explicitly estimate this correlation. To achieve this, another method of assessing dietary fat intake, such as a biomarker, is needed. The error, if any, in this method must not be correlated with any error component in the usual self-reported instruments. The recent progress with biomarkers for macronutrients, such as doubly-labeled water for total energy intake and urinary nitrogen for protein intake, together with research on new biomarkers for assessing total fat or carbohydrate intake, may eventually enable us to measure such correlations and better understand the nature of the biases in self-reports. Until then, our research suggests that measurement error adjustment should be done on a study-specific basis, sensitivity analyses of the type demonstrated in this paper should be conducted, and even then the results of nutritional epidemiologic studies relating diet to disease have to be interpreted cautiously.

ACKNOWLEDGMENTS

Dr. Carroll's research was supported by a grant from the National Cancer Institute (CA-57030) and by the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-E509106).

The authors thank Drs. Carolyn Clifford (the Women's Health Trial Vanguard Study), Walter Willett and Donna Spiegelman (The Nurse's Health Study), Arthur Schatzkin (the National Institutes of Health - American Association of Retired Persons Diet and Health Study), and Elaine Flagg, Ralph Coates, Eugenia Calle, and Michael Thun (the American Cancer Society Prevention II Nutrition Survey Validation and Red Meat Consumption Study), for providing us with the data used in this paper.

REFERENCES


APPENDIX 1

Comparison of Measurement Error Models With and Without Person-Specific Bias

The purpose of this appendix is to show that models, such as that of Spiegelman et al. (28), which do not allow for person-specific bias in a reference instrument lead to a greater attenuation factor compared with the new model proposed in this paper, and therefore that they account for only part of the attenuation if, in fact, this bias takes place. In our notation, the Spiegelman at al. measurement error model can be written as

\[ Q_i = \beta_0 + \beta_1 T_i + r_i \]  

(A.1)

\[ F_{ij} = T_i + e_{ij}, \quad j = 1, \ldots, m_i \]  

(A.2)

\[ W_i = \eta_0 + \eta_1 T_i + v_i \]  

(A.3)

where \( W_i \) denotes a third method of assessing true intake, e.g., a biomarker, with
In the absence of repeat measurements on \( Q \) and \( W \), it is impossible to distinguish between person-specific bias and random errors, so the respective components \( r \) and \( v \) incorporate both these errors. Although this model allows for non-zero correlation between within-person errors \( r \) and \( e_F \), the reference method \( F \) is assumed to have no person-specific bias, so that

\[
\text{cov}(r, v) = \text{cov}(e_F, v) = 0. \tag{A.4}
\]

Assumption A.5 lies at the heart of the difference between this model and model 10-11 introduced in this paper. As follows from equation 4, the attenuation factor in both models is calculated as

\[
\lambda = \frac{\text{cov}(Q, T)}{\text{var}(Q)} = \frac{\beta_1 \sigma_T^2}{\text{var}(Q)}. \tag{A.6}
\]

Since for both models, \( \text{cov}(Q, W) = \beta_1 \eta_1 \sigma_T^2 \) and \( \text{cov}(F, W) = \eta_1 \sigma_F^2 \), we have

\[
\beta_1 = \frac{\text{cov}(Q, W)}{\text{cov}(F, W)}. \tag{A.7}
\]

The expression for \( \sigma_T^2 \), however, depends on assumption A.5. Under this assumption,

\[
\sigma_T^2 = \text{cov}(F_{ij}, F_{ii}), j \neq l,
\]

and it follows from equations A.6–A.7 that the attenuation factor is given as

\[
\lambda^* = \frac{\text{cov}(Q, W)}{\text{cov}(F, W)} \times \frac{\text{cov}(F_{ij}, F_{ii})}{\text{var}(Q)}, j \neq l. \tag{A.8}
\]

In model 10–11, allowing for person-specific bias in the reference instrument \( F \) leads to \( e_{Fij} = s_i + u_{ij} \), so that \( \text{cov}(e_{Fij}, e_{Fii}) = \sigma_s^2, j \neq l \), and \( \sigma_T^2 = \text{cov}(F_{ij}, F_{ii}) - \sigma_s^2 \). As a result, the attenuation factor for model 10–11 is given by

\[
\lambda = \frac{\text{cov}(Q, W)}{\text{cov}(F, W)} \times \frac{\text{cov}(F_{ij}, F_{ii}) - \sigma_s^2}{\text{var}(Q)}, j \neq l,
\]

which, as follows from expression A.8, is always smaller (and hence indicates greater attenuation) than the attenuation factor \( \lambda^* \) calculated using model A.1–A.5. In other words, not allowing for person-specific bias leads to an underestimate of the effect of measurement error.

### APPENDIX 2

**Prentice's Measurement Error Model**

The purpose of this appendix is to show that the model of Prentice (8) imposes the constraint that the correlation between the person-specific biases in the reference instrument and the FFQ is the ratio of their standard deviations. In our notation, for person \( i \), repeat measure \( j \), and body mass index tertile \( v \), Prentice’s measurement error model can be written as

\[
Q_{ijv} = B_{iv} + s_{iv} + e_{ijv}; \tag{A.9}
\]

\[
F_{ijv} = T_{iv} + s_{iv} + u_{ijv},
\]

where it is assumed that both the FFQ and the reference measure share the same person-specific bias \( s \), and, in addition, that the FFQ has a component \( B \) which may depend on the true intake \( T \). It is also assumed that all random variables on the right-hand side of the equations are mutually independent, so that

\[
\text{cov}(B_{iv}, s_{iv}) = 0. \tag{A.10}
\]

Regressing \( B \) on \( T \) leads to

\[
B_{iv} = \beta_0 v + \beta_1 T_{iv} + \delta_{iv}, \tag{A.11}
\]

where, according to assumption A.10,

\[
\text{cov}(\delta_{iv}, s_{iv}) = 0. \tag{A.12}
\]

Substituting expression A.11 into formula A.9, the model for the FFQ can be written as

\[
Q_{ijv} = \beta_0 v + \beta_1 T_{iv} + r_{iv} + e_{ijv},
\]

where

\[
r_{iv} = \delta_{iv} + s_{iv}.
\]

It then follows from condition A.12 that

\[
\text{cov}(r_{iv}, s_{iv}) = \var(s_{iv}),
\]

so that

\[
\rho_{r_{iv}, s_{iv}} = \sigma_{s_{iv}} / \sigma_{r_{iv}},
\]

as claimed.