

COMMENTARY

Dealing With Dietary Measurement Error in Nutritional Cohort Studies

Laurence S. Freedman, Arthur Schatzkin†, Douglas Midthune, Victor Kipnis

Manuscript received October 7, 2010; revised April 20, 2011; accepted April 26, 2011.

Correspondence to: Victor Kipnis, PhD, Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, 6130 Executive Blvd, EPN-3131, MSC 7354, Bethesda, MD 20892-7354 (e-mail: kipnisv@mail.nih.gov).

† Deceased

Dietary measurement error creates serious challenges to reliably discovering new diet–disease associations in nutritional cohort studies. Such error causes substantial underestimation of relative risks and reduction of statistical power for detecting associations. On the basis of data from the Observing Protein and Energy Nutrition Study, we recommend the following approaches to deal with these problems. Regarding data analysis of cohort studies using food-frequency questionnaires, we recommend 1) using energy adjustment for relative risk estimation; 2) reporting estimates adjusted for measurement error along with the usual relative risk estimates, whenever possible (this requires data from a relevant, preferably internal, validation study in which participants report intakes using both the main instrument and a more detailed reference instrument such as a 24-hour recall or multiple-day food record); 3) performing statistical adjustment of relative risks, based on such validation data, if they exist, using univariate (only for energy-adjusted intakes such as densities or residuals) or multivariate regression calibration. We note that whereas unadjusted relative risk estimates are biased toward the null value, statistical significance tests of unadjusted relative risk estimates are approximately valid. Regarding study design, we recommend increasing the sample size to remedy loss of power; however, it is important to understand that this will often be an incomplete solution because the attenuated signal may be too small to distinguish from unmeasured confounding in the model relating disease to reported intake. Future work should be devoted to alleviating the problem of signal attenuation, possibly through the use of improved self-report instruments or by combining dietary biomarkers with self-report instruments.

J Natl Cancer Inst 2011;103:1086–1092

The notion that there is a connection between our diet and our health goes back to biblical times (1). Since the discovery that consumption of citrus fruit protected sailors from developing scurvy (2), many other relationships between diet and disease have been found (3). Nevertheless, for many chronic diseases, the link with dietary intake, if it exists, remains obscure.

Many research designs for studying diet–disease relationships have been used, including animal feeding experiments, migrant studies, ecological epidemiology studies (in which the unit of analysis is a population rather than an individual), and randomized trials, but the two most commonly used are the case–control and cohort study designs. In both studies, participants report their dietary intake using a self-report instrument, usually a food-frequency questionnaire (FFQ). This instrument aims to measure the usual (ie, average) daily intakes of foods and nutrients over the past several months. However, intake estimates that are derived from this instrument invariably differ from the true intake values for several reasons: subjects may find it difficult to recall and average their intakes over the long term, reported intakes may be influenced by psychological factors such as social desirability, and consumption frequencies and average portion sizes of food groups (eg, cold breakfast cereal) may be imperfectly translated to specific nutrient

amounts. Thus, in nutritional epidemiology studies that use self-report instruments, the measured exposure (ie, the estimated intake) has an error that is often substantial and probably larger than that for most other exposures of common epidemiological interest.

Measurement error can be classified into two types: differential and nondifferential. Differential measurement error is the error that is related to the outcome of interest and can occur in a case–control study when case subjects recall their diet with different error than control subjects, resulting in recall bias. This type of measurement error is less likely to occur in a cohort study because diet is usually reported long before the diagnosis of the disease. Here we concentrate on nondifferential measurement error—error that is uncorrelated with disease—and our comments relate only to cohort studies. Measurement error in nutritional case–control studies has not been studied extensively and requires a separate discussion.

Nondifferential measurement error in the measured exposure creates three problems: 1) bias in estimated relative risks; 2) loss of statistical power to detect diet–disease relationships; and 3) in some circumstances, invalidity of the conventional statistical tests for detecting those relationships. We discuss each problem in turn.

How Serious Are These Problems?

Bias in Relative Risks

In univariate disease models that assess associations between disease and a single dietary intake, “classical” measurement error in the exposure attenuates the estimated relative risks (ie, it brings them closer to the null value of 1.0). Classical measurement error is nondifferential additive error that is independent of the true exposure and has mean zero and constant variance. However, dietary measurement error is not usually classical, but instead involves bias that is related to true intake, in addition to random variation (4). The “flattened-slope phenomenon,” in which subjects with a high level of intake tend to underreport their intake and subjects with a low level of intake tend to overreport their intake, inflates the estimated relative risk (5), but the random variation attenuates it. In combination, random variation usually prevails, still leading to overall attenuation of the relative risk estimate (6).

How great is this attenuation? To answer this question, one needs to compare the flawed measurement with an exact measure of usual intake or, in the absence of an exact measure, a “proper reference instrument” (ie, an unbiased measure whose errors are unrelated to usual intake and to errors in the FFQ) (4). Unfortunately, few such measures of either type are available. Other self-report instruments, for example, 24-hour recalls, are biased and their errors are correlated with errors in the FFQ (6). The few measures that are known to be proper reference instruments are “recovery” biomarkers, which have a known quantitative time-associated relation between dietary intake and recovery (excretion) in human waste (7). One recovery biomarker is doubly labeled water for assessment of energy expenditure (8), which, assuming that the study subject is in energy balance, measures energy intake. Other recovery biomarkers are 24-hour urinary nitrogen excretion (9) and 24-hour urinary potassium excretion, which are used to measure intakes of protein and potassium, respectively.

In 1999, the US National Cancer Institute initiated the Observing Protein and Energy Nutrition (OPEN) Study (10), in which 261 male and 223 female adult volunteers completed an FFQ (twice), a 24-hour recall (twice), one doubly labeled water assessment (twice in a subsample of 25 persons), and a 24-hour urinary potassium and urinary nitrogen assessments (twice each). From this study, it was possible to estimate the level of relative risk attenuation when using an FFQ as the main instrument in a cohort study for five exposures: energy, protein, potassium, protein density (the ratio of protein to energy), and potassium density.

Attenuation is quantified by the attenuation factor—a multiplicative factor that operates on the true regression coefficient in the disease model. The smaller the coefficient the greater the attenuation of the relative risk estimate. For energy, the estimated attenuation factor (SE) was 0.08 (0.03) for men and 0.04 (0.03) for women. The attenuation factor (SE) for protein was 0.16 (0.03) and 0.14 (0.04) for men and women, respectively, for potassium, 0.29 (0.04) and 0.23 (0.06), for protein density 0.40 (0.07) and 0.32 (0.08), and for potassium density, 0.49 (0.07) and 0.57 (0.08) respectively (6). These attenuation factors would cause a true relative risk of 2.0 to be estimated, on average, as 1.03–1.06 for energy, 1.10–1.12 for protein, 1.17–1.22 for potassium, 1.25–1.32 for protein density, and 1.40–1.48 for potassium density. Although the

attenuation for the micronutrient potassium (namely, reduction to a relative risk of 1.17–1.22) appeared less extreme than that for the macronutrient protein (reduction to a relative risk of 1.10–1.12), it was nevertheless still substantial. Importantly, energy adjustment improved the attenuation: the attenuation coefficients for protein density and potassium density were larger (and therefore less extreme) than those for absolute protein and potassium. However, by any standard, these results indicate a considerable amount of attenuation that can be attributed to the dietary measurement error in an FFQ, and, if the problem is greatest for macronutrients, it is also present for micronutrients.

Later we show that this concern about attenuation applies also to multivariable models that assess associations of disease with several error-prone dietary exposures.

Loss of Statistical Power

Accompanying the severe degree of attenuation is an equally severe loss of statistical power. Calculations based on data from the OPEN study indicate that to compensate for the loss of statistical power resulting from the use of an FFQ, one would need study samples that are 25–100 times larger for the energy exposure, 10–12 times larger for the protein exposure, and five to eight times larger for protein density. In cohort studies of rare diseases, these sample size inflations necessitate the conduct of enormous studies with hundreds of thousands of participants.

Nutritional epidemiologists have indeed met this challenge by establishing large prospective cohort studies such as the Nurses’ Health Study (11), the European Prospective Investigation into Cancer and Nutrition (12), and the National Institutes of Health—American Association of Retired Persons (NIH–AARP) Diet and Health Study (13), as well as by conducting meta-analyses of cohort studies [eg, (14)]. Later, we discuss whether increasing sample size adequately solves the problem of loss of statistical power.

Invalidity of Conventional Statistical Tests

For a single mismeasured exposure in the disease model, the usual statistical test of the null hypothesis (no exposure effect) remains theoretically valid even though the estimated relative risk is attenuated. However, in multivariable disease models with two or more mismeasured exposures, the validity of conventional statistical tests is no longer guaranteed. In this case, estimated relative risks may become attenuated, inflated, or can even change direction, and consequently, one cannot tell whether a statistically significant relative risk indicates a real association. This concern is important because investigators often include more than one nutritional exposure in disease models. For example, energy adjustment models are commonly used, where a nutrient of interest is included in some form, together with energy (15). Theoretically, the standard, residual, density, and partition energy-adjusted models (16,17) are all subject to this worrying concern.

This change in nature of the bias in the estimated relative risk in models with two or more mismeasured exposures arises from a phenomenon known as residual confounding. When two explanatory variables are in a model and one is mismeasured, then if the variables are correlated, the exactly measured one will “adopt” part of the effect of the mismeasured one. When there are two nutritional intake variables and both are mismeasured, each will adopt a

part of the effect of the other, and the fractions of the effect that are adopted will depend on the relative sizes of the errors and the correlations between them.

To evaluate the extent of this problem, we need to calculate the fraction of the other variable's effect that will be adopted by the exposure of primary interest. We call this fraction the contamination factor. We have used OPEN study data to estimate these contamination factors for protein, energy, and potassium in combination with a variety of other nutrients. Table 1 displays estimated contamination factors for a model with protein density, potassium density, energy, and one other nutrient density, for a selection of nutrients, by sex. The estimated contamination factors are generally small, which means that, at worst, rather small fractions of the effects of one nutritional variable are transferred to another because of residual confounding. Among the 60 estimated values in Table 1, the largest in absolute value is -0.20 , the contamination of the carbohydrates effect from the protein effect. By using the false discovery rate for independent estimates (18) to adjust for multiple comparisons, we found that not one of the 60 estimates is statistically significantly different from zero at the 5% level. (This test provides only a rough guide because the estimates in Table 1 are not mutually independent.)

In summary, the results in Table 1 suggest that residual confounding arising from multiple nutritional variables in the same model does not have a very large impact on relative risk estimates. However, there are two caveats. First, the standard errors of the estimated contamination factors are large (Table 1). Additional information from new validation studies [eg, (19,20)] will allow a more precise estimation of contamination factors. Second, the contamination factors that can be studied are currently limited to

those relating to protein, potassium, energy, and one further nutrient. Contamination between combinations of nutrients that have no recovery biomarkers cannot be studied. Thus, recommendations that are based on such data are founded on seriously incomplete information. However, the estimates in Table 1 represent the best information currently available.

Our summary statement, if true, is somewhat comforting, because it means that usual statistical significance tests of associations between dietary intakes and disease, unadjusted for measurement error, will not be seriously biased. Such tests will not necessarily have adequate statistical power, and so null results should be treated with due caution. However, if a usual (unadjusted) test detects a statistically significant association, then one need not worry that the result is mainly due to residual confounding from another dietary variable in the model. In other words, the results in Table 1 provide good news regarding concern over false-positive results, but no comfort regarding false negatives.

How Best to Deal With These Problems?

Statistical Analysis: Bias in Relative Risks

Given that estimated relative risks are biased because of dietary measurement error, how should researchers analyze and report their results? It seems obvious that the conventional relative risk estimates from the disease risk model (usually a logistic or Cox regression model), albeit biased, should be reported. Also, for FFQ data, the model should include some form of energy adjustment; otherwise, the attenuation is likely to be so severe as to preclude useful results. A more difficult issue is whether to report estimates

Table 1. Contamination factors of selected nutrient and food densities with energy, protein density, and potassium density, estimated from Observing Protein and Energy Nutrition Study data*

Nutrient or food density†	Sex	Protein density (SE)	Potassium density (SE)	Energy (SE)
Energy	Male	-0.01 (0.03)	0.13 (0.05)	NA
	Female	0.03 (0.05)	0.10 (0.06)	NA
Protein	Male	NA	-0.01 (0.09)	0.08 (0.05)
	Female	NA	0.00 (0.10)	0.06 (0.05)
Potassium	Male	-0.05 (0.06)	NA	0.04 (0.04)
	Female	0.00 (0.07)	NA	-0.04 (0.05)
Total fat	Male	-0.03 (0.07)	0.00 (0.08)	0.05 (0.05)
	Female	-0.02 (0.08)	-0.08 (0.10)	-0.07 (0.05)
Saturated fat	Male	-0.03 (0.05)	-0.04 (0.07)	0.10 (0.04)
	Female	-0.01 (0.06)	-0.07 (0.08)	-0.02 (0.04)
Carbohydrates	Male	-0.20 (0.09)	-0.12 (0.12)	-0.06 (0.07)
	Female	-0.07 (0.12)	-0.19 (0.15)	0.14 (0.08)
Fiber	Male	-0.05 (0.06)	0.08 (0.08)	-0.12 (0.04)
	Female	0.10 (0.06)	0.12 (0.08)	0.00 (0.04)
Vitamin C	Male	0.00 (0.03)	0.03 (0.04)	-0.02 (0.02)
	Female	0.09 (0.04)	0.05 (0.05)	-0.03 (0.03)
Vitamin E	Male	0.04 (0.05)	0.07 (0.06)	-0.05 (0.03)
	Female	0.03 (0.07)	0.14 (0.08)	0.00 (0.04)
Sodium	Male	0.04 (0.10)	0.10 (0.13)	-0.17 (0.07)
	Female	0.08 (0.12)	-0.06 (0.15)	0.01 (0.08)
Fruits and vegetables	Male	0.00 (0.04)	0.09 (0.06)	-0.03 (0.03)
	Female	0.13 (0.05)	0.06 (0.07)	-0.04 (0.03)

* Contamination factors for the first three nutrients or food densities (energy, protein, and potassium) are for a disease model with the following three explanatory variables: food-frequency questionnaire (FFQ) protein density, FFQ potassium density, and FFQ energy. For the remaining nutrients or food densities, the contamination factors are for a disease model with four explanatory variables: FFQ protein density, potassium density, energy, and the selected nutrient density. NA = not applicable.

† All nutritional variables were log-transformed.

that are adjusted, in some way, for measurement error. The method of regression calibration (21,22) has often been used to provide such adjustment. However, implementation of this method is not straightforward, as we now explain.

Implementing regression calibration requires knowledge of the relationship between the true value and the observed value, as do other methods of statistical adjustment for measurement error. Such knowledge can be obtained from a specially conducted validation study, in which a random subsample of participants in the main study complete a proper reference instrument in addition to the FFQ. The reference instrument usually used in the validation study is a more detailed self-report, such as a multiple-day food record or 24-hour recall. Unfortunately, these instruments do not meet the requirements for a proper reference instrument because their measurement errors are related to true intake and are correlated with those of the FFQ (6,23). The question therefore arises: Should researchers perform regression calibration with an imperfect reference instrument, such as a 24-hour recall, and report the resulting adjusted relative risk estimates, or should they report only unadjusted relative risks together with a statement that the estimates are biased, most probably attenuated? We examined this question by using data from the OPEN study to compare attenuation and contamination factors based on a 24-hour recall reference instrument with those based on a biomarker reference instrument. We first explain the relevance of this comparison to the stated question.

A commonly used form of regression calibration adjustment that was proposed by Rosner et al. (21) can be implemented by first estimating the attenuation and contamination factors on an appropriate scale in which the reference instrument and FFQ are approximately linearly related, and then applying these factors to the unadjusted relative risk estimates to calculate the adjusted relative risks. Thus, using OPEN data, we calculated the attenuation and contamination factors using a 24-hour recall reference instrument and compared them with the values obtained using recovery biomarkers. If the factors calculated using a 24-hour recall appear similar to those calculated using recovery biomarkers, then the adjusted risk estimates based on regression calibration using a 24-hour recall reference instrument will also be similar to those that would have been obtained had a proper reference instrument been used. Table 2 and Supplementary Table 1 (available online) present such a comparison for attenuation and contamination factors, respectively.

Table 2 shows that the 24-hour recall-based attenuation factors for energy are overestimated compared with those based on recovery biomarkers; however, for protein density and potassium density, the differences between the 24-hour recall-based and recovery biomarker-based attenuation factors are smaller than those for energy, and there is no overall trend toward larger values using the 24-hour recall as reference. For protein density, the differences between biomarker-based and 24-hour recall-based estimates are in different directions for men and women (although these differences are not nominally statistically significant and could be chance fluctuations); for potassium density, the differences between biomarker-based and 24-hour recall-based estimates are small.

The 24-hour recall-based contamination factors tend to have higher absolute values compared with recovery biomarker-based contamination factors (Supplementary Table 1, available online). This tendency is not very marked; for example, the mean absolute

Table 2. Attenuation factors for energy, protein density, and potassium density estimated using recovery biomarkers vs using a 24-hour recall as the reference instrument, results from the Observing Protein and Energy Nutrition Study

Nutrient*	Sex	Recovery biomarker based† (SE)	24-h Recall based‡ (SE)
Energy	Male	0.08 (0.03)	0.21 (0.04)
	Female	0.03 (0.03)	0.09 (0.05)
Protein density	Male	0.43 (0.07)	0.35 (0.07)
	Female	0.33 (0.08)	0.45 (0.06)
Potassium density	Male	0.57 (0.08)	0.59 (0.05)
	Female	0.61 (0.08)	0.62 (0.07)

* All nutritional variables were log-transformed.

† Based on a multivariable disease model with food-frequency questionnaire (FFQ) energy, protein density, and potassium density as explanatory variables. Note that the values for energy and protein density are not identical to the attenuation factors for these variables that are reported in the text. This is because the values cited in the text were those for a univariate model.

‡ Based on a multivariate disease model with FFQ energy, protein density, and potassium density as explanatory variables.

value for 24-hour recall-based factors is 0.07 and the mean absolute value for biomarker-based factors is 0.06. However, 13 of 60 24-hour recall-based factors are nominally statistically significant at the 5% level, and the three factors with the lowest *P* values—protein density with potassium density for women, carbohydrate density with protein density for men, and energy with potassium density for men—are statistically significant at the 5% level even after the false discovery rate correction. It therefore appears that use of the 24-hour recall as the reference instrument can occasionally lead to seriously inflated estimates of the contamination factor.

How should we translate into practical recommendations the similarities and differences between 24-hour recall-based factors and recovery biomarker-based factors shown in Table 2 and Supplementary Table 1 (available online)? We have conducted a series of calculations that postulate scenarios involving different combinations of true relative risks, and to those scenarios we applied the 24-hour recall-based factors to evaluate the size of bias expected in the estimates of these relative risks. We compared three possible approaches to relative risk estimation: 1) no adjustment for measurement error; 2) a univariate measurement error adjustment for each relative risk estimate; and 3) a multivariate measurement error adjustment for the set of relative risks. In the univariate adjustment method, we ignored contamination factors, assuming them to be close to zero. This assumption is motivated by the small values in Table 1 and because the 24-hour recall sometimes misestimates a contamination factor (Supplementary Table 1, available online). The univariate measurement error adjustment is simple, requiring only division of the unadjusted relative risk estimate by the 24-hour recall-based attenuation factor for that variable. The multivariate adjustment method uses the full set of 24-hour recall-based attenuation and contamination factors, as described by Rosner et al. (21). We asked whether either adjustment method, based on a 24-hour recall reference, reduces the bias in estimated relative risks compared with no adjustment, and if so, which adjustment method performs better. (See the Supplementary Material [available online] for a description of the design and results of these investigations.) From the results

obtained, we conclude that both the univariate and multivariate methods, on average, improve estimation compared with no adjustment. Although the univariate method performed somewhat better, on average, than the multivariate method for men, there was little difference between the methods for women. We repeat the caveat that these calculations were limited to models that include protein, potassium, and energy.

As explained in more detail in the Supplementary Material (available online), our current recommendation for nutritional cohort studies that use an FFQ as the main instrument and have an internal validation study with a self-report reference instrument (eg, a 24-hour recall or multiple-day food record) is to use either the univariate or multivariate method of adjustment in reporting relative risks. We expect that most investigators would prefer using the simpler univariate method, which, on the basis of current evidence, is acceptable. Note, however, that our recommendation refers to energy-adjusted intake variables used in the density or residual models. Univariate adjustment for the unadjusted intakes used in the standard and partition models is inappropriate because the attenuation factor for the nutrient would be too small; the multivariate adjustment is recommended in this case.

Note that in implementing the regression calibration adjustments that we are recommending, the attenuation and contamination factors should be estimated from the validation study after adjustment for the (exactly measured) confounders included in the disease model. For example, for the univariate method, the attenuation factor is estimated as the linear slope of the reference instrument value on the FFQ value in a multiple regression that also includes the confounders. Other practical issues regarding the regression calibration adjustment, including notes on the design of appropriate validation studies and the special design and analysis requirements that arise when the nutrient or food of interest is consumed seasonally or episodically, may be found at the end of the Supplementary Material (available online).

For cohort studies that do not include an internal validation study, investigators should try to obtain information from validation studies of other cohorts that are conducted in a population similar to the study population of interest and with a similar main study instrument. The measurement error adjustment should be based on this external information.

Study Design: Loss of Power and Signal Attenuation

We have shown that regression calibration can alleviate the biased estimation caused by dietary measurement error, but unfortunately, it usually does not recover the lost statistical power. In fact, when using the univariate adjustment, the ratio of the adjusted risk estimate to its standard error will be somewhat smaller than that for the unadjusted estimate, and the *P* value, larger. Assuming small contamination effects, this is also likely to be true for the multivariate adjustment method.

The classical remedy to a loss of statistical power is to increase the study sample size. As mentioned earlier, several very large nutritional observational cohort studies have been conducted (11–13). A recent report from one of those studies, the NIH–AARP cohort study (24), illustrates this approach to recovering lost statistical power. The study report included 188 736 postmenopausal women, 3501 of whom were diagnosed with breast cancer

during follow-up. The estimated energy-adjusted hazard ratio for breast cancer for the highest vs lowest quintile of percent energy from total fat was 1.11 (95% confidence interval [CI] = 1.00 to 1.24) and the test for trend across quintiles was statistically significant ($P_{\text{trend}} = .017$). However, as the authors note in their discussion, unmeasured or incompletely ascertained confounders could have influenced the results.

The NIH–AARP report illustrates a general principle: When relative risk attenuation due to measurement error is severe, the usual remedy of increasing sample size does not necessarily solve the problems of interpretation caused by measurement error. As demonstrated in the OPEN study, dietary measurement error can cause attenuation of sizable relative risks to observed values of 1.25 or less. When the unadjusted relative risk is this low, it becomes uncertain whether the observed association, even if statistically significant, is due to the exposure or to unmeasured confounders (see Supplementary Material [available online] for more details). With observed associations this weak, the effects of unmeasured confounders can become dominant, and, unlike random variation, these effects cannot be removed by increasing the sample size. This problem arises directly from the “signal attenuation” that is familiar to nutritional epidemiologists. Increasing the sample size can recover the loss of power, but will not lessen the attenuation of the signal.

Another common approach for dealing with low statistical power is to conduct a meta-analysis of available cohort studies as has been done in studying the association between dietary fat intake and breast cancer (14,25,26). However, several investigators have cautioned against overinterpreting apparently highly precise results reported from meta-analyses of observational studies, for reasons similar to the “signal attenuation” argument elucidated above. For example, Egger et al. (27) warn that confounding can distort findings from observational studies and of the consequent “danger that meta-analyses of observational data produce very precise but equally spurious results.” They conclude that “the statistical combination of data should not therefore be a prominent component of reviews of observational studies.”

Nutritional epidemiologists must therefore turn their attention to addressing signal attenuation as well as low statistical power. Several steps toward this aim are already being taken. First, new self-report instruments are being developed that could be used in large studies and that may have improved measurement characteristics over those of the FFQ. Three studies (28–30) suggest some improvement in the ability to detect a disease–diet association when a more detailed instrument is used compared with the FFQ. In these studies, statistically significant associations between dietary intakes and disease were found using 7-day diaries or multiple-day food records, whereas the association as measured through an FFQ did not achieve statistical significance. Previously, the barrier to using such instruments in large studies has been the labor intensive and costly coding of the records. Now, automated versions of a 24-hour recall (31,32) promise to overcome that barrier, and pilot studies of their use are in progress.

Schatzkin et al. (33) used data from OPEN to evaluate the potential gain from the use of a 24-hour recall as the main instrument in a cohort study. They estimated, using mathematical modeling for four repeats of a 24-hour recall, attenuation factors

for protein density of 0.50 (SE = 0.09) for men and 0.40 (SE = 0.13) for women compared with values for a single FFQ of 0.40 (SE = 0.07) for men and 0.32 (SE = 0.08) for women. Thus, for energy-adjusted components, although one should not expect dramatic improvements in attenuation and statistical power from the use of multiple repeats of an automated 24-hour recall, there is reason to hope that there will be some worthwhile gains over the use of an FFQ. Moreover, if new cohort studies were designed to include repeat 24-hour recalls plus an FFQ determination, further improvements in attenuation may be seen from combining information from the two instruments. A method for combining such information has been described for intakes of a single nutrient or a single food that is “episodically consumed” (ie, not typically consumed every day by all in the population) (34). This method now requires extension to models with multiple nutrients and foods.

A second way of addressing the signal attenuation is to combine information from self-report instruments with measurements of dietary biomarkers. We have already mentioned the few recovery biomarkers that have been used to study the measurement error in self-report instruments. However, there is a much larger class of dietary biomarkers—the so-called “concentration” biomarkers (7)—which are known to be correlated with dietary intakes of different foods or nutrients, although they do not represent the exact level of intake. These include serum carotenoids, lipids, and vitamins. Freedman et al. (35) have proposed combining measurements of these biomarkers with information from self-reports to strengthen the signal and increase statistical power in analyses of diet–disease relationships. They have recently demonstrated this approach in an analysis of the association between dietary lutein and zeaxanthin intakes and nuclear cataracts (36). In their example, the estimated odds ratio for disease using the FFQ was 0.77 compared with an estimated odds ratio of 0.68 using the combined FFQ–biomarker measure. This difference in odds ratios represented an increase of 50% in the signal (the log odds ratio increased in absolute value from 0.26 to 0.39).

With a similar objective of incorporating biomarkers into dietary assessments, Prentice et al. (37) have proposed a large feeding study in which participants consume their usual diet, and an array of biomarkers is measured at study baseline and conclusion. The data that accrue may allow a rational combination of a wider range of biomarkers with information from self-reports, potentially further advancing progress in combating the research problems caused by dietary measurement error.

Concluding Remarks

In summary, we have provided recommendations for statistical analysis to deal with the biased estimation of relative risks that arises from dietary measurement error. In addition, with regard to study design, we have highlighted the dual problems of loss of statistical power and signal attenuation. The statistical analysis recommendations are based on data from OPEN. When information from recently completed and currently active validation studies using recovery biomarkers becomes available, our recommendations can be checked and, if necessary, updated. Signal attenuation represents

the major obstacle to progress, and thus the emphasis of future work should be on alleviating it.

References

- Lilienfeld AM. Ceteris paribus: the evolution of the clinical trial. *Bull Hist Med.* 1982;56(1):1–18.
- Lind J A. Treatise of the Scurvy. In Three Parts. Containing an Inquiry into the Nature, Causes and Cure, of that Disease. Together with a Critical and Chronological View of what has been Published on the Subject. Edinburgh: Sands, Murray and Cochran for A Kincaid and A Donaldson; 1753.
- Scrimshaw NS, SanGiovanni JP. Synergism of nutrition, infection and immunity: an overview. *Am J Clin Nutr.* 1997;66(2):464S–477S.
- Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. *Am J Epidemiol.* 1999;150(6):642–651.
- Wacholder S. When measurement errors correlate with truth: surprising effects of nondifferential misclassification. *Epidemiology.* 1995;6(2):157–161.
- Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol.* 2003;158(1):14–21.
- Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr.* 2002;5(6A):969–976.
- Schoeller DA. Measurement error of energy expenditure in free-living humans by using doubly labeled water. *J Nutr.* 1988;118(11):1278–1289.
- Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *Am J Clin Nutr.* 1985;42(6):1276–1289.
- Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol.* 2003;158(1):1–13.
- Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses’ health study. *Am J Nurs.* 1978;78(6):1039–1040.
- Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol.* 1992;3(10):783–791.
- Schatzkin A, Subar AF, Thompson FE, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health–American Association of Retired Persons Diet and Health Study. *Am J Epidemiol.* 2001;154(12):1119–1125.
- Smith-Warner SA, Spiegelman D, Adami HO, et al. Types of dietary fat and breast cancer: a pooled analysis of cohort studies. *Int J Cancer.* 2001;92(5):767–774.
- Willett WC, Stampfer MJ. Total energy intake; implications for epidemiologic analyses. *Am J Epidemiol.* 1986;124(1):17–27.
- Wilkens L, Lee J. Nutritional epidemiology. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics.* Chichester, UK: Wiley; 1998:3099–3177.
- Kipnis V, Freedman LS, Brown CC, Hartman A, Schatzkin A, Wacholder S. Interpretation of energy adjustment models for nutritional epidemiology. *Am J Epidemiol.* 1993;137(12):1376–1380.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995;57(1):289–300.
- Moshfegh AJ, Rhodes DG, Baer DJ, et al. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *Am J Clin Nutr.* 2008;88(2):324–332.
- Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women’s Health Initiative. *Am J Epidemiol.* 2008;167(10):1247–1259.
- Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol.* 1990;132(4):734–745.
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in non-linear models: a modern perspective. 2nd ed. Boca Raton, FL: Chapman and Hall/CDC; 2006.

23. Martin LJ, Su W, Jones PJ, Lockwood GA, Tritchler DL, Boyd NF. Comparison of energy intakes determined by food records and doubly labeled water in women participating in a dietary-intervention trial. *Am J Clin Nutr*. 1996;63(4):483–490.
24. Thiebaut A, Kipnis V, Chang SC, et al. Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health–AARP Diet and Health Study cohort. *J Natl Cancer Inst*. 2007;99(6):451–462.
25. Hunter DJ, Spiegelman D, Adami H-O, et al. Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. *N Engl J Med*. 1996;334(6):356–361.
26. Boyd NF, Stone J, Vogt KN, Connelly BS, Martin LJ, Minkin S. Dietary fat and breast cancer risk revisited: a meta-analysis of the published literature. *Br J Cancer*. 2003;89(9):1672–1685.
27. Egger M, Schneider M, Davey-Smith G. Meta-analysis spurious precision? Meta-analysis of observational studies. *Br Med J*. 1998;316(7125):140–144.
28. Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet*. 2003;362(9379):212–214.
29. Freedman LS, Potischman N, Kipnis V, et al. A comparison of two dietary instruments for evaluating the fat–breast cancer relationship. *Int J Epidemiol*. 2006;35(4):1011–1021.
30. Dahm CC, Keogh RH, Spencer EA, et al. Dietary fiber and colorectal cancer risk: a nested case-control study using food diaries. *J Natl Cancer Inst*. 2010;102(9):614–626.
31. Subar AF, Thompson FE, Potischman N, et al. Formative research of a quick list for an automated self-administered 24-hour dietary recall. *J Am Diet Assn*. 2007;107(6):1002–1007.
32. Subar AF, Crafts J, Zimmerman TP, et al. Assessment of the accuracy of portion size reports using computer-based food photography aids in the development of an automated self-administered 24-hour recall (ASA24). *J Am Diet Assn*. 2010;110(1):55–64.
33. Schatzkin A, Kipnis V, Carroll RJ, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based OPEN study. *Int J Epidemiol*. 2003;32(6):1054–1062.
34. Kipnis V, Midthune D, Buckman DW, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*. 2009;65(4):1003–1010.
35. Freedman LS, Kipnis V, Schatzkin A, Tasevska N, Potischman N. Can we use biomarkers in combination with self-reports to strengthen the analysis of nutritional epidemiologic studies? *Epidemiol Perspect Innov*. 2010;7(1):2.
36. Freedman LS, Tasevska N, Kipnis V, et al. Gains in statistical power from using a dietary biomarker in combination with self-reported intake to strengthen the analysis of a diet-disease relationship—an example from CAREDS. *Am J Epidemiol*. 2010;172(7):836–842.
37. Prentice RL, Huang Y, Tinker LF, Beresford SAA, Lampe JW, Neuhouser ML. Statistical aspects of the use of biomarkers in nutritional epidemiology research. *Stat Biosci*. 2009;1(1):112–123.

Funding

This work was supported by a contract held by Information Management Services, Inc, with the National Cancer Institute (to LSF); the remaining authors were supported by their institution, the National Cancer Institute at the National Institutes of Health.

Notes

The study sponsors had no role in study design, analysis, or interpretation of the data; in the decision to submit the article for publication; or the writing of the article.

The authors have full responsibility for all ideas in this article.

We thank Amy Subar, Sharon Kirkpatrick, and Susan Krebs-Smith for helpful suggestions.

Affiliations of authors: Biostatistics Unit, Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer, Israel (LSF); Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics (AS) and Biometry Research Group, Division of Cancer Prevention (DM, VK), US National Cancer Institute, Bethesda, MD.