

## Letters to the Editor

### Dietary fat underreporting and risk estimation

Sir,

We would like to express concerns about Drs Heitmann and Lissner's conclusion that associations between fat intake and disease risk in observational studies may be overestimated, rather than attenuated, due to underreporting<sup>1</sup>. Unfortunately, their claim is based on considering underreporting as the only source of measurement error in dietary assessment and seems to be unlikely in practice, at least for macronutrients. Yet it could impact the interpretation of numerous epidemiological studies, notably in the context of the fat and breast cancer controversy<sup>2,3</sup>.

Weaknesses in the demonstration include the objective measurements used to assess underreporting and the lack of consideration of other sources of measurement error. Underreporting in fat intake was indirectly evaluated using reference measurements for protein and energy intakes. Because underreporting was greater in energy than in protein intake, overreporting in percentage energy from protein was assumed to be exactly compensated for by underreporting in percentage energy from fat (fat density). When the relationship between fat density and low-density lipoprotein cholesterol (LDL-C) was then analysed, the correction for underreporting decreased rather than increased the magnitude of the association. We believe that such a result could be anticipated and does not provide a sufficient proof for fat and disease associations being overestimated.

Let  $Q_E$  and  $Q_P$  denote reported energy and protein intake, respectively (as estimated from a diet history questionnaire describing food intake in the previous month), and  $R_E$  and  $R_P$  the corresponding reference measurements ('measured energy and protein intakes' in the authors' words). The reference protein intake was assessed using nitrogen excretion in a single 24-hour urine sample. However, urinary nitrogen measurements are subject to substantial within-subject variability, with a coefficient of variation as large as 13 to 24%<sup>4</sup>. In lieu of a direct 'recovery' biomarker (i.e. doubly labelled water)<sup>5</sup>, 24-hour energy expenditure derived from self-reported physical activity level and basal energy expenditure served as the reference energy intake. In addition to the potential misclassification of physical activity levels, the estimated 24-hour energy expenditure encompasses uncertainties related to the coefficients for body fat and fat-free mass in the equation of basal energy expenditure (respective standard errors 3.9% and 24.3% of the corresponding estimated parameters)<sup>6</sup>, as well as uncertainties related to the equation of body fat as a function of sex, age,

measured impedance, height and weight ( $R^2 = 0.90$ )<sup>7</sup>. Even limited error in the reference energy intake may lead to substantial error in protein density  $R_P/R_E$  which therefore needs to be taken into account.

Systematic error (underreporting) in fat density  $Q_D$  was adjusted for assuming the following equation for the reference fat density:  $R_D = Q_D + \delta$ , where  $\delta = (Q_P/Q_E) - (R_P/R_E) > 0$  denotes the overreporting in protein density. However, even if systematic bias in  $Q_D$  is perfectly corrected,  $R_D$  should include at least within-person random variation still remaining in  $Q_D$ , plus additional random error due to the fact that the correction relies on single-day measurements and estimated components. Thus, at best,  $R_D$  contains classical measurement error and can be represented as  $R_D = T_D + \xi$ , where  $T_D$  denotes true (unmeasured) fat density with variance  $\sigma_T^2$ , and  $\xi$  denotes random error independent of  $T_D$  with mean of 0 and variance  $\sigma_\xi^2$ . Under this model, the slope of the linear regression of LDL-C on  $R_D$  will be attenuated by a factor

$$\lambda_R = \rho^2(R_D, T_D) = \frac{1}{1 + \sigma_\xi^2/\sigma_T^2},$$

where  $\rho(\phi, \varphi)$  denotes the correlation coefficient between random variables  $\phi$  and  $\varphi$ . As for the reported fat density  $Q_D$ , a reasonable measurement error model would be  $Q_D = \beta_0 + \beta_1 T_D + \varepsilon$ , where  $\varepsilon$  is the sum of within-person random error and person-specific bias<sup>8</sup>, has mean of 0 and variance  $\sigma_\varepsilon^2$ . Compared with truth again, the slope of the regression of LDL-C on  $Q_D$  will be attenuated by a factor

$$\lambda_Q = \frac{1}{\beta_1} \rho^2(Q_D, T_D) = \frac{\beta_1}{\beta_1^2 + \sigma_\varepsilon^2/\sigma_T^2}.$$

The observed findings are consistent with  $\lambda_Q > \lambda_R$  rather than  $\lambda_Q > 1$  as concluded in the article<sup>1</sup>. The former inequality suggests that within-person variation  $\sigma_\xi^2$  in  $R_D$  is large compared with between-person variation of true fat density  $\sigma_T^2$ . This should not be surprising, given that part of the variation  $\sigma_\xi^2$  is due to substantial random errors in the components of  $R_D$ ,  $Q_D$  and  $\delta$ , as mentioned above. But this fact does not mean that  $\lambda_Q > 1$  or, equivalently, that  $\rho^2(Q_D, T_D) > \beta_1$ . To the contrary, results of the OPEN (Observing Protein and Energy Nutrition) study with repeated recovery biomarker measurements for protein (24-hour urinary nitrogen) and total energy (doubly labelled water) intakes suggest that this is not the case for protein as well as non-protein density reported on the questionnaire<sup>9</sup>. In OPEN, the attenuation factors were estimated as 0.404 and 0.316 for protein density in men and women respectively, 0.370 and 0.290 for non-protein density<sup>9</sup>.

The case when both the predicted variable (LDL-C) and the predictor variable (fat density) are categorical was also considered and yielded similar results, although not statistically significant<sup>1</sup>. The use of the dichotomised version of LDL-C instead of its continuous version would not change the conclusion because, to an excellent approximation, attenuation factors under logistic regression (for dichotomous predicted variable) are the same as under linear regression<sup>10</sup>. As for a predictor variable categorised into quantiles, the observed relationship can only be attenuated<sup>11</sup>.

According to the intuitive explanation provided in the paper, underreporting was considered to be at least the major, if not the only, part of error in dietary questionnaire measurements. If it were so, then  $\lambda_Q$  would be close to  $1/\beta_1$  and, because usually  $\beta_1 < 1$  due to the flattened-slope phenomenon (high consumers tend to underreport whereas low consumers tend to overreport),  $\lambda_Q$  would indeed be greater than 1. However, empirical data suggest that one cannot neglect random variation in dietary self-report, as it seems in practice to compensate for, and even overwhelm, the overestimating impact of systematic error<sup>9</sup>.

Anne CM Thiébaud<sup>1\*</sup> and Victor Kipnis<sup>2</sup>

<sup>1</sup>National Cancer Institute  
Nutritional Epidemiology Branch  
Division of Cancer Epidemiology and Genetics  
6120 Executive Blvd, Executive Plaza South 3033  
Bethesda, MD 20892, USA

\*Email: thiebauta@mail.nih.gov

<sup>2</sup>National Cancer Institute  
Biometry Research Group  
Division of Cancer Prevention  
6130 Executive Blvd, Executive Plaza North 3124  
Bethesda, MD 20892, USA

DOI: 10.1017/S1368980007339049

## References

- 1 Heitmann BL, Lissner L. Can adverse effects of dietary fat intake be overestimated as a consequence of dietary fat underreporting? *Public Health Nutrition* 2005; **8**: 1322–7.
- 2 Prentice RL, Caan B, Chlebowski RT, Patterson R, Kuller LH, Ockene JK, *et al.* Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. *Journal of the American Medical Association* 2006; **295**: 629–42.
- 3 Freedman LS, Potischman N, Kipnis V, Midthune D, Schatzkin A, Thompson FE, *et al.* A comparison of two dietary instruments for evaluating the fat–breast cancer relationship. *International Journal of Epidemiology* 2006; **35**: 1011–21.
- 4 Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. *Journal of Nutrition* 2003; **133**(Suppl. 3): 921S–4S.

- 5 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 Nutrition* 2002; **5**: 969–76.
- 6 Garby L, Garrow JS, Jorgensen B, Lammert O, Madsen K, Sorensen P, *et al.* Relation between energy expenditure and body composition in man: specific energy expenditure *in vivo* of fat and fat-free tissue. *European Journal of Clinical Nutrition* 1988; **42**: 301–5.
- 7 Heitmann BL. Prediction of body water and fat in adult Danes from measurement of electrical impedance. A validation study. *International Journal of Obesity* 1990; **14**: 789–802.
- 8 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. *American Journal of Epidemiology* 1999; **150**: 642–51.
- 9 Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, *et al.* Structure of dietary measurement error: results of the OPEN biomarker study. *American Journal of Epidemiology* 2003; **158**: 14–21.
- 10 Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* 1989; **8**: 1051–69.
- 11 Kipnis V, Izmirlian G. The impact of categorization of continuous exposure measured with error. *American Journal of Epidemiology* 2002; **155**: S28.

## The authors reply in plain English

Sir,

We would like to thank Drs Thiébaud and Kipnis for their thoughtful comments. We agree that random reporting error is always present and results in attenuated associations. It is clearly necessary to recognise all types of error and their potential impact on epidemiological associations.

As the commentary correctly points out, we were not focusing on random error in our paper<sup>1</sup>. Even if in some cases random error may overwhelm biases from systematic error, this may not always be the case. Whether true associations are overestimated or underestimated depends on the magnitudes of these two types of error, as well as on the direction of the bias in relation to the underlying association. The purpose of our paper was to illustrate, not to prove, that non-random error can in theory inflate an association.

The effects of non-random errors on diet–disease associations are not always appreciated. For instance, in a recent re-analysis of data from the OPEN (Observing Protein and Energy Nutrition) study by a team including ourselves, Dr Kipnis and other researchers from National Cancer Institute<sup>2</sup>, we concluded that obesity-related reporting errors require much further investigation. Although the OPEN data are indeed unique in being able to characterise both types of error, it must be kept in mind that they are based on a highly selected study sample. Clearly, we need to improve our knowledge about person-specific and other non-random errors, as well as our ability to communicate about them.