Appendix

Derived reference measures based on the observed biomarkers

In the OPEN study, the replications of the DLW measurement were available in only a small sample of 25 individuals (14 men and 11 women). This fact did not affect the results for total energy intake since the DLW measurements were remarkably consistent across replications. The coefficient of variation in the DLW measurements was only 5.1%, in effect indicating that energy expenditure was measured with very little error.

However, a technical difficulty arose in the analysis of energy-adjusted protein. The error in the biomarker-based derived reference measure was almost entirely influenced by the error in the UN measurements where the coefficient of variation was 17.6%. As a result, attempting to estimate the within-person variance of the derived reference measure as a parameter in the model led to relatively large standard errors in the main analysis and to instability in the procedure for bootstrap calculations.

Based on these facts, in dealing with the derived reference measurements for energy-adjusted protein, we used the following convention. When defining biomarker-based reference measures for nutrient density and nutrient residual, we used the first DLW observation with both the first and second repeat UN observations. In theory, this induced some correlation between repeat biomarker-based reference observations and therefore lead to somewhat underestimated within-person variation, but the measurement error in DLW was so small that this underestimation could be ignored in practice. For example, because the logarithm of the ratio of UN to DLW is equal to the difference between their corresponding logarithms, using the first DLW measurement underestimates within-person variation in the derived reference measure by within-person variance of DLW. From the components-of-variance analysis of the 25 subjects with two DLW observations, this variation was estimated as 0.0025. Subtracting this value from the estimates of between-person variation of true protein density intake (Table 2) would decrease them by approximately 8%. This would have very little effect on estimated model parameters that depend on between-person variation of true intake. For example, it would increase the estimated correlation coefficient between reported and true protein density by 4%.

Commentary: An OPEN assessment of dietary measurement errors

Martyn Plummer and Rudolf Kaaks

In epidemiological studies of chronic disease risk in relation to diet, a crucial question is whether assessments of dietary intake can accurately characterize an individual's habitual intake of foods and nutrients. Over the last two decades, this question has been addressed in numerous validation studies. The OPEN study provides the best answer yet. It is a true landmark study both because of its size and the thoroughness of its design.

The main innovation of the OPEN study is the use of doubly labelled water (i.e. water made from stable isotopes of both hydrogen and oxygen) to measure energy expenditure. Doubly labelled water is extremely expensive to produce. It is therefore unlikely that the OPEN study will be replicated in the foreseeable future. With this lack of reproducibility in mind, we would like to consider the extent to which the conclusions of the OPEN study can be generalized.

Before considering the conclusions of OPEN, it is worth reviewing the evolution of dietary validation studies. Initially, dietary measurement validation studies were based on a comparison of two assessment methods, one of which (often based on a series of weighed food consumption records) was assumed to provide a perfectly valid intake measurement. This strong validity assumption was later relaxed by the use of statistical models for measurement error. These models impose certain constraints on the design of validation studies, where the nature of the constraint depends on the purpose of the study. If the aim is to correct for the 'attenuation' effect of measurement error on estimates of disease risk then two independent measurements are required, one of which must be unbiased. If the aim is to completely characterize the error properties of the dietary assessment methods then three independent measurements are required.

In practice, three main categories of dietary assessment can be distinguished: questionnaires for assessment of habitual, long-term intake; methods based on recording of actual food intake on one or more days (e.g. weighed food records, 24-hour diet recall interviews), and biomarkers of diet. It is clear that the measurement errors of instruments in the first two categories are correlated, not least because the same food tables must be used when converting foods to nutrients. This leaves only certain biomarkers as a possible alternative. One set of biomarkers of particular interest consists of those based on the urinary recovery of chemical substances from diet. Such recovery-based markers allow the computation of absolute daily intakes of nutrients in
time. They also have the advantage that repeat measurements may be assumed independent, if the repeats are sufficiently widely spaced, so that the necessary statistical assumptions for validation studies may be fulfilled. Recovery-based markers are currently available for energy expenditure, and for protein, sodium, and potassium intake. The first two of these biomarkers were used in the OPEN study. A previous validation study by Day et al. used biomarkers for protein, sodium, and potassium.

There are a number of interesting conclusions from the OPEN study. First, the food frequency questionnaire (FFQ) was found to have very low validity for both total energy and protein intake measurements: attenuation factors were found to be close to 0.1 in both cases. It is disconcerting to observe that the FFQ measurements predict so little variation in protein intake levels. A second, unexpected, result is that a single 24-hour recall (24-HR) appeared to predict both protein and energy intakes better than the FFQ. This is somewhat surprising given the well-documented day-to-day variation in dietary intakes. A third interesting feature of the OPEN study is the comparison between the results for absolute and energy adjusted protein intake, measured by the protein density. An energy-adjusted analysis simulates an intervention in which total energy intake is fixed, but the composition of the diet changes. Two motivations for this analysis have been put forward: Firstly, energy requirements depend on body size, physical activity, and metabolic efficiency of each individual, which collectively may confound relationships of absolute nutrient intakes with disease risk; adjustments for total energy intake might reduce such confounding effects. Secondly, it is claimed that measurements of absolute intake contain a substantial degree of measurement error that may be eliminated by energy adjustment. This latter claim is only partially borne out by the OPEN study. Whereas energy adjustment substantially decreases the degree of attenuation for protein intake measured by the FFQ, it has a more modest effect on the correlation with true intake. This suggests that the main problem of the FFQ is its inability to give absolute estimates of protein intake on a well calibrated scale. For the 24HR, energy adjustment has very little effect on the attenuation factor, and no clear effect on correlation. There is even a decrease in correlation among women after energy adjustment.

A general limitation of validation studies is that the results are context specific. The results of a validation study are not necessarily transferable to another population, or even to other nutrients in the same population, because the summary statistics (attenuation factor and correlation coefficient) depend on both the characteristics of the instruments used and the heterogeneity of intake in the population. This point, which is acknowledged by the authors, creates a paradox. On the one hand, the OPEN study has set a high standard for validation studies which is unlikely to be matched. On the other hand, the extent to which the results can be generalized is not clear. Some evidence for generalizability to other populations comes from the general agreement with the result of the previous validation study by Day et al.5 When it comes to generalization to other nutrients, however, there are no previous studies with a similar design to compare with, because all such studies are limited to recovery-based biomarkers. We note that the day-to-day variation in protein and total energy intakes is low compared with that of many other nutrients. For other nutrients, for which intake levels vary more over time, but also more between individuals, the FFQ may perform much better, both in absolute terms and relative to a 24HR. Having said this, we are not quite as optimistic as the authors that new biomarkers can be developed to measure other nutrients, so our speculation will probably remain unchallenged by empirical evidence.

References