Wirfält et al. Respond to Drs. Block and Willett

Elisabet Wirfält,¹ Bo Gullberg,² and Robert W. Jeffery³

We appreciate the lengthy responses provided by Drs. Block (1) and Willett (2) and agree with them that a large number of studies have demonstrated the utility and robustness of food frequency questionnaires (FFQs). We also agree that frequency intervals currently are preferable to exact frequencies on self-administered FFQs in large population surveys. However, in smaller surveys using interview-administered instruments, more information will be obtained with exact frequency response alternatives. The current study suggests that the degree of detail when estimating frequencies and/or portion sizes may be of different importance for the ranking ability of different foods and nutrients. If the number of foods listed on a questionnaire were the only feature distinguishing intakes, then the longer instrument should have shown stronger correlations with the reference method on all accounts. In addition, cognitive research has indicated that frequency judgments of individuals are reasonably accurate and reflect a sensitivity to the rates of consumption. However, frequency judgments between individuals have much lower accuracy, probably due to lack of a common standard between study participants (3). Moreover, the response scale presented on a questionnaire will serve as a reference for the respondent when behavior frequencies are judged (4). This implies that food-frequency scales need to mirror common consumption frequencies in the population. Food-specific frequency scales, for instance, may calibrate frequency responses between individuals better than scales with the same frequency intervals for all foods.

The study was not primarily designed to test the two FFQs, which undoubtedly is a weakness. However, participants were entering a lengthy weight reduction study and, at the time of the baseline examinations, should have had high motivation to provide good quality data. Unclear responses were carefully controlled by the study coordinator, either in person or by telephone. Learning effects due to the order of administration of FFQs should have been to the benefit of the Willett instrument. However, it is possible that participants may have been more impatient with the second of two long food lists, resulting in less-accurate responses. The observed findings cannot merely be disregarded as "statistical noise," and our interpretation is not speculative. The Bonferroni procedure is a conservative way to control for overall error rates in multiple comparisons. Its assumption—that the comparisons are totally independent of one another—is not fulfilled in this study, and true significant differences could, therefore, have been underestimated. Additionally, an overall pattern in the performance of the two instruments emerged—one instrument suggested higher correlations for nutrients contained in a few nutrient dense foods, and the other suggested higher correlations for nutrients contained in many foods.

It should not surprise anyone that the Block and Willett FFQs overall provided diet reports more similar to each other than to diet recalls. The structures of list-based methods are similar, the cognitive processing required by the respondent is similar, and, therefore, the measurement errors are also expected to be similar (5). As recently discussed at the Third International Dietary Assessment Method Conference (6), the random variation, or statistical noise, observed in dietary assessment is not only due to random variation in intake, but could arise from instrument biases and characteristics of individuals (i.e., inappropriate memory cues; difficulties of the individual to provide the requested frequency or portion-size information; or reporting biases related to cultural beliefs, behaviors, or other characteristics of the respondent). Two components of bias have been described: person-specific biases, which are systematic within individuals but appear random across individuals, and group-specific biases, which are systematic across individuals (7). Energy adjustment, commonly used in nutrition epidemiology studies, does not always take care of observed biases and errors of dietary instruments (8). In addition, energy adjustment may be inappropriate, es-
especially if both the examined nutrient (i.e., fat) and energy are risk factors for the disease (5). Researchers concerned with improving the accuracy in usual diet assessment therefore need to identify the specific instrument designs that better reflect dietary intakes and to define the relations between observed biases and personal characteristics (indicators of obesity, physical activity, sociodemographics, social desirability measures, etc.). Instruments need to be compared with standards that have a minimum of correlated errors (i.e., biomarkers) (9). When the error structures of observed diet reports are better understood, we will also have the means to control for them in analysis.

REFERENCES