Nut-11. Quantitative variables: explain categorization of dietary and nutritional data (e.g., use of N tiles and handling of nonconsumers) and the choice of reference category, if applicable.

Example 1. “We combined FFQ items to create variables reflecting intakes of 1) total sugary beverages (combining sugar-sweetened soft drinks, fruit juice, and fruit drinks), 2) sugar-sweetened soft drinks (high-sugar carbonated beverages, such as cola), and 3) artificially sweetened soft drinks (sugar-free carbonated beverages, such as diet cola). We created new intake categories to ensure that an adequate number of participants were retained in each intake group across each variable. Cut points were determined before conducting the main analyses based on the relative distribution of intake for each variable. Total sugary beverage consumption was examined as <1/d (reference), 1–2/d, and >2/d; sugar-sweetened soft drink intake was examined as 0/wk (reference), ≤3/wk, and >3/wk; and artificially sweetened soft drink intake was examined as 0/wk (reference), ≤6/wk, and ≥ 1/d” .

Explanation.

In nutritional epidemiology, nutrient and food variables are often examined in categories delineated by N-tiles (e.g., quintiles cutoffs indicating fifths of the distribution; see also Nut-14). This is one way of handling outliers, exaggerated intakes (i.e., potential measurement errors), and nonconsumption. Nonconsumption is common in certain foods (e.g., meat) and in alcohol. The design features of dietary assessment tools may result in exaggerated reports of high intakes. For instance, if many different types of a food item (e.g., fish) are listed in an FFQ this may result in a misleadingly inflated intake in absolute terms (see also Nut-8.1). The true intakes of those individuals who report very high intakes may, however, correctly belong to the higher end of the distribution. In addition, foods with a high concentration of certain nutrients (e.g., vitamin A) may be consumed episodically and unequally in the population, potentially resulting in skewed distributions. Categorization of exposure variables is also needed when a specific cutoff has been recognized, and intakes below or above certain levels need to be compared (e.g., to express the compliance with dietary recommendations). A clear description of the selected categories and cutoffs, the mean or median values of categories, the reference category, and how nonconsumers were handled will be helpful to readers.

In studies that estimate disease risks, the preferred reference category should be one that is stable and includes a sufficient number of study subjects. Although the reference category is often the category with the lowest (or highest) nutrient intake, there may be particular reasons for selecting another category. For instance, individuals who report zero consumption of alcohol may be a mix of those who have never tasted alcohol and those who previously consumed large amounts and recently stopped. In such cases, a more suitable reference may be regular consumers of low amounts. Similarly, a midcategory of the intake distribution might be chosen as the reference category when both high and low intakes are proposed to be associated with the outcome

(i.e., U-shaped association). Excluding nonconsumers in analysis could be informative both in descriptive and etiologic studies but could also bias the findings. In association studies, nonconsumers may serve as reference category for RR estimates, or be the measure of interest. That is, nonconsumers may be maintained in the sample for population mean estimates, or excluded when the average portion size is estimated. If nonconsumers are excluded, their key characteristics should be reported and compared with those of the examined study sample to ensure clarity when interpreting study findings.