ELSEVIER

Journal of Food Composition and Analysis 21 (2008) S20–S26

JOURNAL OF FOOD COMPOSITION AND ANALYSIS

www.elsevier.com/locate/jfca

Study Review

Development of food composition databases for food frequency questionnaires (FFQ)

Suzanne McNutt*, Thea P. Zimmerman, Stephen G. Hull

Westat, 3949 E. Viewcrest Drive, Salt Lake City, UT 84124, USA

Received 6 November 2006; received in revised form 15 May 2007; accepted 31 May 2007

Abstract

The food frequency questionnaire (FFQ) is the most commonly used dietary assessment instrument for large epidemiological studies because it collects "usual" diet information at a low cost to researchers, and imposes a low burden to survey participants. The FFQ consists of a list of foods with little descriptive detail, and the respondent answers questions about the frequency of each food on the list. A FFQ food composition database links each food in the list to its associated food component values. The steps required to develop a food composition database for a FFQ are to define the purpose of the study, identify the source(s) of data, create a food list, make decisions about the need for portion size information, and incorporate the food component values into the database. The sources of data most often used in the United States are the US Department of Agriculture (USDA) Nutrient Database for Standard Reference or the USDA Food and Nutrient Database for Dietary Studies. Rather than develop a new FFQ for a study or survey, researchers may evaluate current, validated FFQs to determine if they can be used.

Keywords: Dietary database; Food frequency questionnaires; FFQ; Dietary assessment; Food composition

1. Introduction

Researchers use three primary methods to collect food intake information. Cross-sectional evaluations are the most common type of dietary assessment, and 24-h recalls (24-HRs) interviews and food records (FRs) are typically used to collect these data. Retrospective assessments, which evaluate past diet to determine associations with present disease states such as osteoporosis, heart disease, or cancer, generally use food frequency questionnaires (FFQs). Prospective assessments and intervention studies, used to evaluate dietary change and compare to physiological endpoints, use all three methods. Although FFQs provides less specificity and greater measurement error compared to 24-HRs and FRs, they are the most commonly used instruments in large dietary epidemiology studies because they collect "usual" diet at a low cost to researchers, take less time than the other methods, and have a low respondent burden.

Dietary assessment instruments are linked to food composition databases to estimate intakes of food components (e.g. nutrients, alcohol, fiber, and flavonoids). The databases should be accurate, up-to-date, and contain the foods and food components of interest; be specific, precise, and uniform; and have no missing values. The US Department of Agriculture (USDA) Nutrient Database for Standard Reference (SR) (USDA, 2006b) is the most commonly used food composition database in the United States (Thompson and Subar, 2001). It includes information on more than 140 food components for 7000 foods. The information in the SR is used to develop the USDA Food and Nutrient Database for Dietary Studies (FNDDS) (USDA, 2006a) which is used for dietary assessment in the National Health and Nutrition Examination Surveys (NHANES).

A survey database can be used to evaluate food component intakes from foods reported on 24-HRs and FRs because the information collected is usually detailed

*Corresponding author. Tel.: +1 801 453 9268; fax: +1 801 453 9269.
E-mail address: susiemcnutt@westat.com (S. McNutt).

0889-1575/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jfca.2007.05.007

enough for a specific food code to be assigned to the food consumed. However, FFQs consist of a list of fairly generic foods; thus each food line item corresponds to many variations of that food in a survey database. For example, "pizza" is a typical FFQ line item, but the food components associated with that food represent the numerous varieties of pizza that people eat. A FFQ database is constructed by linking the foods in the FFQ list to a food code that represents specific food components. The FFQ database used might be developed with data from the SR or FNDDS and is tailored to meet the needs of the dietary study.

The purpose of this paper is to describe the typical steps in the development of FFQ databases and to outline the questions researchers should ask when choosing FFQs for research studies.

2. Developing a FFQ and associated database

Developing a FFQ and database is a highly technical task that requires attention to many details. There are five general steps; however, researchers designing new instruments often borrow from existing FFQs and may only need to complete a few of the steps. Whether starting from scratch or altering an existing FFQ, investigators should not undertake this task without careful consideration.

2.1. Step 1: Define the FFQ purpose

The first step is to define the purpose of the FFQ. The population of interest may be adults, pregnant women, school-aged children, or some other specific group. The intent may be to collect information on the total daily diet, only certain food groups such as fruits and vegetables, or foods that contain specific food components such as calcium and carotenoids. The objective of the data collection may be to rank individuals (to discriminate according to intake) or to provide a measure of estimated intake. FFQs designed to estimate intakes must collect comprehensive information on the diet which results in a longer questionnaire. Researchers interested in estimating intakes must use more detailed FFQs than researchers who want to evaluate food groups and specific food components or to rank individuals.

2.2. Step 2: Identify source(s) of information

The second step is to identify the source(s) of information needed to build the FFQ database. The foods should represent those commonly consumed in the population of interest. Information about the foods consumed can be obtained from national survey data such as the National Health and Nutrition Examination Survey (NHANES) (NCHS, 1999–2004) and the New Zealand 2002 National Children's Nutrition Survey (Ministry of Health, 2003); or from cohorts that include the specific population of interest such as the Harvard Nurse's Health Study (Willett, 1998a, b) and the NIH-AARP Diet and Health Study (Schatzkin et al., 2001). Food composition data can be obtained from databases such as the SR (USDA, 2006b) or FNDDS (USDA, 2006a), from provisional tables on the food contents of specific food components (USDA, 2006c), or from a combination of sources (Jensen et al., 2004).

2.3. Step 3: Construct the FFQ food list

There are two methods for building a FFQ food list: one is to identify the foods consumed by the target population, and the other is to use stepwise multiple regression techniques to identify foods that are the best predictors of dietary factors of interest. In the first approach, food consumption information is evaluated to determine which foods are most often consumed by the target population. The data may come from national surveys or from studies of specific populations. The multiple regression approach is favored when the goal is to identify foods that are most predictive of certain dietary components such as percent energy from fat or intakes of fiber, calcium, or vitamin C. A list of foods is compiled from a food composition database that represents important food component sources; pilot tests are conducted to test the predictive value of the foods in the list; and foods with low predictive values are discarded. After multiple iterations of analysis a modest number of foods are identified that are most predictive of the food components of interest (Willett, 1998a, b). Regardless of the method used to construct the food list, it should be tested with the target population to make sure the food names and descriptions are understandable and that it provides the type of information sought by the investigators.

2.4. Step 4: Assess the need for portion sizes and methods of assessing them

There are three options with regard to portion sizes in FFQs: (1) collect no portion information; (2) specify a portion size as part of the food line item; or (3) include discrete questions for portion sizes for each food.

2.4.1. No portion information collected

An FFQ that does not collect information on portion size is called a "non-quantitative" FFQ. The Block Questionnaire for Kids 2–7 (Fig. 1), used in epidemiological research, is an example of this option (Block, 2004). An age-specific portion size is applied to each food line item that represents the average amount consumed by that population based on national survey data. This may be a good option if the estimated intake of the total diet is not required and the population of interest is homogeneous (e.g. young children) such that the researcher believes that standard portions are representative of the amounts consumed in that population. In addition, if the researcher feels it is unlikely the respondent can answer the portion questions, as when parents are reporting their pre-school