Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Food Science 2 (2013) 52 – 59

Procedia Food Science

36th National Nutrient Databank Conference

ARS, USDA updates food sampling strategies to keep pace with demographic shifts

Pamela Pehrsson*, Charles Perry, and Marlon Daniel

Nutrient Data Laboratory,Beltsville Human Nutrition Research Center, Agricultural Research Service, 10300 Baltimore Avenue, Beltsville, MD 20705

Abstract

The National Food and Nutrient Analysis Program (NFNAP) was implemented in 1997 as a collaborative food composition research effort between USDA and NIH. The goal of this program is to obtain nationally representative estimates of the nutritional components of important foods consumed in the US for inclusion in the USDA National Nutrient Databank System; to date, analytical food composition data generated for over 1800 foods have vastly improved overall data quality in the database. The NFNAP sampling approach was updated in 2001 using 2000 US Census data and recently updated to use 2010 Census population estimates. This design, like the 2001 design, employs a three-stage, stratified, probability-proportional-to-size (PPS) sample selection process; 1) county selection (based on population density); 2) supermarket outlets within selected counties (based on annual sales); and 3) specific brands of foods (based on market share data). In the first stage, Census regions (4), divisions and states were used to obtain a self-weighting sample of population centres, ensuring geographic dispersion across the 48 conterminous states; 48 locations were selected, with nested subsets of 24, 12 and 6 locations. Due to demographic changes in the population and congressional redistricting it was necessary to revise the sampling scheme to reflect these changes. With the increased penetration of warehouse-type retail outlets into the grocery industry, the sampling frame must be adjusted to include these purchase locations. Food samples which are collected nationally according to a statistically rigorous sampling approach are consistent with national representativeness and allow better estimates of the mean and variability than convenience sampling or less rigorous options.

Selection and peer-review under responsibility of National Nutrient Databank Conference Steering Committee

Keywords: National sampling; food composition

1. Introduction

The US Department of Agriculture's (USDA) Nutrient Data Laboratory (NDL) develops high quality food composition databases for foods available in the US food supply. This paper describes the second revision of the National Food and Nutrient Analysis Program (NFNAP) sampling plan, implemented in 2012, for the national collection of food samples from retail outlets for nutrient analysis. A more detailed history of this program, implemented in 1997, and applications of the sampling approach and NFNAP in general are presented in earlier NDL publications (1,2,3,4). The basic objectives of NFNAP are to secure reliable estimates with known variability for the nutrient content of food and beverages consumed by the US population.

The initial and subsequent updated sampling plans are based on a stratified three-stage design using the most current population density data from the US Bureau of the Census and food sales data for retail

* Corresponding author. Tel.: +1-301-504-0693; fax: 1-301-504-0692.
E-mail address: pamela.pehrsson@ars.usda.gov

Selection and peer-review under responsibility of National Nutrient Databank Conference Steering Committee
doi:10.1016/j.profoo.2013.04.009

outlets in selected locations and product market shares, both from ACNielsen, Inc. Selection of locations (population density), retail outlets (sales), and specific brands (market share data) are selected probability-proportional-to-size, so that any county, store or brand in these three selection levels has a chance of being selected; the greater the proportion to the total, the greater the probability of being selected. The primary focus of this paper will be the first stage, selection of locations. The sampling plan provides a self-weighting nationally representative sample set of the food products.

2. Methods

2.1. Chromy's PMRPPS Procedure

Chromy's algorithm, a probability minimum replacement (PMR) probability proportional to size (PPS) sampling scheme, was again used to select a stratified sample of counties to purchase foods for nutrient analysis (5,6,7). A sequential sampling scheme considers a frame's sampling unit in a predefined order. PMR sample designs are PPS designs which allow some sampling units to be selected more than once. Chromy's procedure identifies the following elements:

n(i) = number of times unit i is selected in sample

n = sample size

S(i) = size measure for sample unit i

S(+) = sum of size measures for all units in frame

q(i) = E[n(i) = nS(i)/ S(+)]

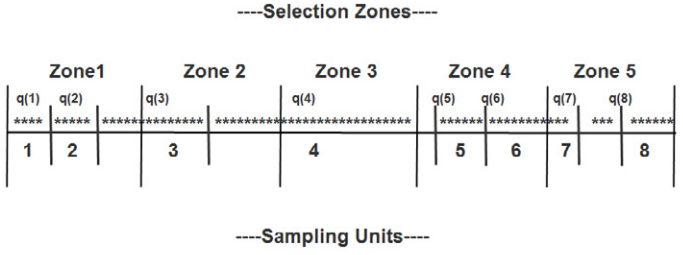

Fig.1.Chromy's PMRPPS Sampling Procedure Designed to Choose 5 Sampling Zones from Among 8 Population Units (Cities or Counties)

This procedure divides the ordered frame into n zones of size $S(+)/n$. One sampling unit is selected PPS from each zone and each unit I has a line segment of length $q(i)$ associated with it; each line segment either falls entirely within one sampling zone or overlaps two or more zones. Figure 1 illustrates an example where a sample of size five is to be drawn from eight available sampling units. If $q(i)$ is greater than one, then sampling unit i completely covers one or more zones and is considered a self-representing unit (unit 4 in Figure 1). These units will appear in the sample at least one time. If a unit is in part of two adjoining zones but is not self-representing, it can be selected in one of the two sampling zones but not both (units 3 and 6 in Figure 1). When a single unit is selected within each zone, the sample is stratified by the ordering of the frame; the ordering considers control variables highly correlated with the quantity being measured so that conterminous units are similar. In other words, the variance is reduced as long as units in close proximity are more homogenous than units in the population at large.

2.2. Objectives for County Selections

In the 2012 design using 2010 data similar to previous designs, the selected counties are not only geographically dispersed across the nation and regions but are statistically representative with respect to