



United States  
Department of  
Agriculture

Center for Nutrition  
Policy and Promotion

CNPP-22

May 2008

# Development of the CNPP Prices Database



# Development of the CNPP Prices Database

**Andrea Carlson<sup>1</sup>**

**Mark Lino<sup>1</sup>**

**WenYen Juan<sup>1</sup>**

**Kristin Marcoe<sup>1</sup>**

**Lisa Bente<sup>1</sup>**

**Hazel A. B. Hiza<sup>1</sup>**

**Patricia M. Guenther<sup>1</sup>**

**Ephraim Leibtag<sup>2</sup>**

**<sup>1</sup>U.S. Department of Agriculture, Center for Nutrition Policy and Promotion**

**<sup>2</sup>U.S. Department of Agriculture, Economic Research Service**

## **Acknowledgment**

The authors gratefully acknowledge the expert assistance of Dr. Shanthy Bowman, USDA Agricultural Research Service, in developing the methodology used to create the CNPP Prices Database.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410 or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

**May 2008**

## Introduction and Background

When people make decisions about what to eat, they may consider such factors as taste, ease of preparation, nutrition, what others enjoy eating, and price. Many researchers seek to understand what makes some people choose healthful diets while others choose less healthful diets. To conduct these studies, researchers must understand as much of the context of the decisions as possible: What foods are normally enjoyed by an individual and his or her co-diners?<sup>1</sup> How much time does an individual or household have to prepare food? Are there health conditions or other factors that might prompt an individual to value health and nutrition more so than others do? What are an individual's or household's budget restrictions?

To conduct these studies, researchers need data to describe those factors that might influence the eating decision. The National Health and Nutrition Examination Survey (NHANES) provides a wealth of information on the dietary intake, medical history, and current health status of its respondents. The 2007-08 cycle of NHANES also includes a module with questions addressing consumer behavior, which will allow a deeper understanding of the food choices when these data are available. However, the National Research Council points out in its report, *Improving Data to Analyze Food and Nutrition Policies*, that there are limiting factors when using NHANES for economic analysis. For example, income and asset data as well as time resources that might affect food preparation are not included. Detailed information on neighborhood characteristics and specific levels of participation in food assistance programs are also missing, as are price data (National Research Council, 2005).

The U.S. Department of Agriculture's (USDA) Center for Nutrition Policy and Promotion (CNPP) has taken the first step in attaching price data to NHANES dietary recall data. CNPP has completed an estimation of the expenditures on foods that were reported in the NHANES 2001-02 as having been consumed by participants. This estimate uses scanner data and assumes all food is prepared at home, but uses convenience products available in the grocery store. CNPP is working with the USDA's Economic Research Service (ERS) to develop the next phase of attaching price data to NHANES dietary recall data: estimating the prices of food eaten away from home, as well as estimating regional and seasonal variation for some foods. This first step—attaching price data for home-prepared foods to NHANES dietary recall data—has three main goals:

1. To attach price data to dietary intake data in order to update the USDA Food Plans (discussed later) and to provide a basis for other price- and nutrient-related research.
2. To attach nutrient information to products reported as purchased in the Nielsen Homescan™ consumer panel to provide more information to researchers using these panel data to examine household food purchase behavior.
3. To develop an automated process that will allow for updates as new dietary intake, food composition, and price data become available.

---

<sup>1</sup>If an individual wishes to dine with others, he or she may not always be able to choose his/her personal favorites all the time because the preferences of others may influence the foods prepared or the choice of restaurant.

The last goal is important because these data are updated regularly. USDA receives price and food composition data annually and dietary intake data every other year. In addition, other researchers could use the process to create prices for their own dietary recall data. This report will focus mainly on the first goal, but the other goals will be briefly discussed.

This release of the price data is the version used in estimating the most recent release of the USDA Food Plans and meets the specific needs of the USDA Food Plans. The methodology was originally developed by Shanthy Bowman for the 1999 Thrifty Food Plan (TFP) (Bowman, 1997). To understand the 2001-02 CNPP Prices Database, it is important to understand how the USDA Food Plans are used and why the data needs may differ from other research projects.

USDA estimates the cost of consuming a nutritious diet, as defined by the 2005 Dietary Guidelines for Americans (U.S. Department of Health and Human Services & USDA, 2005), at four expenditure levels: the Thrifty Food Plan (TFP) and the Low-Cost, Moderate-Cost, and Liberal Food Plans. Each plan is a representative and healthful diet and provides a cost estimate for food. The TFP cost level serves as the basis for the maximum food stamp allotment, and the market basket produced from the TFP is used to adjust the allotment for inflation, using the Consumer Price Indexes for food. The Low-Cost Plan is used by bankruptcy courts to set food budgets and by the U.S. Department of State to set a food allowance for visiting students. The U.S. Department of Defense uses the Liberal Plan to set the Basic Allowance for Subsistence rate for all servicemembers. The Low-Cost, Moderate-Cost, and Liberal Plans are used by divorce courts to set alimony and by the CNPP report *Expenditures on Children by Families* (Lino, 2007), which is used to set State-level child support guidelines and foster care payments. Because the maximum Food Stamp Program allotment does not vary among the 48 coterminous States, a single set of national average prices must be used in the TFP to estimate the cost of food. In addition, under the regulations of the Food Stamp Program, hot food cannot be purchased and food stamps cannot be used in restaurants. The Basic Allowance for Subsistence is also meant to cover food prepared at home or in the communal dining facility, not food purchased in restaurants. For more information on the Thrifty Food Plan, please see, *Thrifty Food Plan, 2006* (Carlson, Lino, Juan, Hanson, & Basiotis, 2007). Information on the other three plans can be found in the report *The Low-Cost, Moderate-Cost, and Liberal Food Plans, 2007* (Carlson, Lino, & Fungwe, 2007).

In summary, this release of the prices database was developed for estimating the USDA Food Plans for the 48 coterminous States. As a result, these prices assume one national average price for each food that was reported as having been consumed. We also assume that foods are prepared at home, although convenience foods are considered, including those used to reduce time in food preparation, such as bottled sauces, packaged mixes, canned soups, and frozen foods (including frozen fruits and vegetables). It should be noted that no individual faces national average prices, and unlike some information associated with consumer choice, the price is readily available as the consumer is putting food into the grocery cart. Economic theory suggests that variation in price will have a much larger effect on choices than

information that is not readily available to the consumer. For example, there are differences in nutrient quantity between two tomatoes that appear to be identical to the average food shopper. To arrive at the amount of vitamin A in a tomato, several tomatoes are tested and an average is considered. Researchers may be tempted to assume this average vitamin A content could be treated the same as the average price. It cannot, because most consumers are not aware of the differences, and those that are aware cannot tell in the grocery store which tomato is better. However, the consumer can determine which tomato costs less and will recognize that there is variation in the price of tomatoes. Even with these limitations, these data do allow researchers to include prices in an analysis of NHANES data.

## Data and Methods

### Data

The CNPP Prices Database was developed by combining information from four different data sets: NHANES 2001-02 (Centers for Disease Control [CDC] & USDA, 2004), the USDA Food and Nutrient Database for Dietary Studies version 1.0 (FNDDS) (USDA, Agricultural Research Service [ARS], Food Surveys Research Group, 2004), the National Nutrient Database for Standard Reference (Release 16.1 and 18) (USDA, ARS, 2004, 2005), and the Nielsen Homescan™ Consumer Panel (Nielsen, 2005). More details on each data set are provided later. Briefly, CNPP used NHANES data to create a list of foods that were reported by Americans as having been consumed. These foods are in the as-consumed form. The FNDDS has recipes that break these foods down into ingredients and provides details on the overall moisture and fat losses and gains from the cooking process. However, not all ingredients in the recipe files are in the as-purchased form. To calculate the number of grams in the as-purchased form, we used the National Nutrient Database for Standard Reference, which provides information to calculate the moisture losses and gains caused by the cooking of individual ingredients and the percentage of the purchased weight that is considered to be refuse loss. Finally, we calculated the cost of the ingredients by taking an average of the prices paid by consumers who participated in the Nielsen Homescan™ Consumer Panel.

*NHANES.* The CNPP Prices Database was developed using the 24-hour dietary recall data in NHANES, which were the most recent consumption data available at the time this project began (CDC & USDA, 2004). The foods reported as having been consumed are in the as-consumed form. In a typical 2-year cycle of NHANES, there are about 5,000 to 6,000 foods that are reported as having been consumed. In developing the prices database, we used the NHANES data to develop a list of as-consumed foods. The NHANES is conducted by the National Centers for Disease Control and Prevention through the National Center for Health Statistics. Data are released on a biannual basis and include extensive medical and family history data, but as discussed previously, economic and demographic variables are limited, making economic analysis difficult. The sample size is approximately 10,000 individuals for each 2-year cycle.

***FNDDS.*** The FNDDS (USDA, ARS, Food Surveys Research Group, 2004) is created by ARS to calculate the nutrient values of foods, based on the nutrient content of ingredients used to prepare the foods. The FNDDS includes all as-consumed foods reported in NHANES, along with the quantity of each ingredient required to prepare the food. The data set also contains the moisture losses and gains from cooking the dish once all ingredients are assembled, such as the moisture lost from baking lasagna or toasting bread. Based on the NHANES list of foods reported as having been consumed, CNPP created a list of approximately 2,200 ingredients (both basic and convenience items) that would require a price, plus about 1,600 foods such as breads, cereals, crackers, soup, and frozen foods that could be purchased in the ready-to-eat or ready-to-heat form. FNDDS version 1.0 was developed for NHANES 2001-02 and was used in the creation of prices associated with these data; however, FNDDS 2.0 has been released and will be used in estimating the prices for NHANES 2003-04 data, using the semi-automated process developed. FNDDS 2.0 reflects changes in food processing and fortification that occurred since the development of version 1.0. Recipes and weights of fast-foods and frozen foods also reflect portion sizes sold in 2003-04.

***National Nutrient Database for Standard Reference.*** The National Nutrient Database for Standard Reference (USDA, ARS, 2004, 2005) is maintained by the USDA ARS Nutrient Data Laboratory. This is the major source of food composition data in the United States, and it contains nutrient data for over 7,000 foods. The Standard Reference files also contain the moisture content, the refuse loss, and the gram weights of standard quantities for foods that are not sold by weight (i.e., generally sold by fluid ounces or quantity, such as packaged fruits or baked goods).

***Nielsen Homescan™ Consumer Panel.*** The Nielsen Homescan™ consumer panel data (2001, 2002) were used to calculate national average retail food prices (Nielsen, 2005). Because this is not a publicly available data set, more information on the data is provided in this report. Homescan™ households are recruited based on their demographic information to ensure sufficient representation for demographic variables such as household income, family composition, education, and geographic location. Each household is equipped with an electronic home-scanning unit, and household members are expected to record every food purchase they make by scanning in the appropriate codes of the food products that they purchase for home consumption. For products with a Universal Product Code (UPC), the participant scans the UPC bar code. Nielsen provides the participating household with a book of bar codes for products that do not contain UPCs, such as in-store bakery products, bulk goods, and many produce items. Homescan™ data are unique in that panelists record food purchases across all outlet channels, including grocery, drug, mass-merchandise, club, supercenter, and convenience stores.

Standard demographic information is collected annually from each household, and each household's home market or city and census region are identified for stratification purposes.

Each household is then assigned a projection factor (weight) based on its demographics<sup>2</sup> so that the data can be aggregated to be representative at the market, regional, and national levels. The information captured on a transaction level includes date of purchase, store name and type of store (convenience, small grocery, large grocery, warehouse, discount store), store department identifier, item description, number of units purchased, price paid, and promotions/sales/coupons used (if any). In addition, product attributes, such as the form, brand name, package size, or flavor, are included for most products. CNPP used the item description (product module) and the product attributes to match the ingredients and commercially prepared items to the appropriate products in the Homescan data. To calculate the prices, CNPP used the package size to convert the purchase amount to grams, the number of units purchased, and total price paid as well as the projection factor weight. The price paid reflects a store promotion or sale price available to all consumers who shop in the store. We did not use the coupon discount because coupons are not available to all consumers.

For 2001 and 2002, there were about 8,500 households per year that recorded both their UPC-coded transactions and their random-weight (non-UPC coded) food purchases. This sample was obtained from USDA ERS and used to estimate the price of all foods purchased for home consumption. More information on how many and what types of products are used is provided in the methods section below.

Some researchers question the quality of household panel data when they try to reconcile it with store-level scanner data. There is a perception that the volumetric data from each source should be the same. However, consumer panel data and store data are not always equal because measurement methodologies differ. Store-level scanners record millions of shopping transactions, whereas consumer panel data represent a specific group of shoppers. In addition, panel data represent household-based purchases only; small business or institutional purchases are excluded. Panel data were used in this report to capture all store choices made by households across demographic groups and to allow calculation of all foods including random weight (non-UPC) products that are not usually tracked in store-level scanner data such as produce, some meats, and in-store deli and bakery items. The panel also includes purchases made in discount stores, smaller stores, and independent grocery stores that are not normally included in scanner data.

---

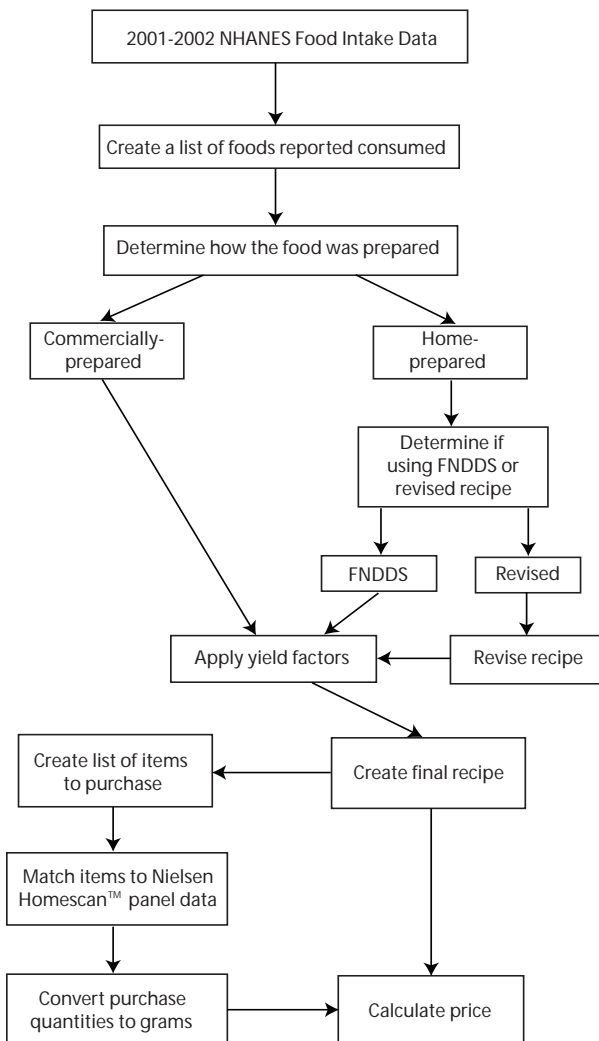
<sup>2</sup>Age, gender, education, occupation of head(s) of household, number of household members, household income, household composition, race, and ethnicity.

## Methods

We followed the methodology developed by Shanthy Bowman to estimate prices for the 1999 TFP (Bowman, 1997). The major steps used in creating the CNPP Prices Database are summarized in figure 1. They are the following:

1. Create a list of foods reported as having been consumed in NHANES and determine how the foods were prepared.
2. Apply yield factors to the individual ingredients and the entire dish.
3. Create the final recipes.
4. Create list of items to purchase.
5. Match the items identified in step 4 to Nielsen Homescan™ Consumer Panel data.
6. Convert as-purchased amounts to grams.
7. Calculate the price.

**Figure 1. Steps To Calculate the CNPP Prices Database**





### **1. Create the List of Foods and Determine How the Foods Were Prepared.**

We used the 24-hour dietary recall data in NHANES 2001-02 to create a list of all foods reported as having been consumed by individual participants. Because the Food Plans do not include alcohol, we excluded alcohol consumed as a beverage.

We next determined which foods were likely to have been purchased as a commercially prepared item and which were likely to have been prepared at home. For foods prepared at home, CNPP staff also determined whether convenience ingredients were used in the preparation. In an effort to incorporate more prepared foods into the Thrifty Food Plan, CNPP staff examined a list of the 50 foods most commonly consumed by low-income individuals from each of the nine major categories of foods in NHANES: dairy; meat, poultry and fish; eggs; beans and nuts; grains; fruits; vegetables; and beverages, fats, and sweets. The list of foods suggested that while cost may be a factor for low-income households, many foods were chosen because of taste and ease of preparation. Many of the frequently consumed items were assumed to be purchased as ready-to-heat foods, such as frozen pizza and canned beans, particularly when CNPP staff believed the time intensity in preparing the frequently consumed foods did not mesh with analysis of time available for low-income consumers to prepare food (Mancino & Newman, 2007). Of the 4,152 foods reported as having been consumed, 322 were changed to ready-to-heat items, rather than multiple ingredient recipes. This number does not include the many foods that were already in the FNDDS as a single, purchasable food, such as breads, fruit, and frozen dinners. For foods that were not purchased as ready-to-heat, CNPP selected common commercially prepared ingredients such as frozen pie crusts; mixes for stove-top side dishes and skillet meals; seasoning packages; canned beans; canned and frozen ready-made sauces and broths; and canned and frozen fruits, vegetables, and soups. The decision on what to include was again based on frequency of consumption and common short-cuts used in preparation.

Next, CNPP reviewed the recipes in the FNDDS. This led to further additions to the list of convenience items used in the recipes. Because the FNDDS is used primarily for nutrient content calculations, some ingredients are broken down more finely than they are found in the purchased form. We also determined that some recipes contained ingredients, such as industrial shortening, dried vegetables, high fructose corn syrup, or ascorbic acid, that indicated the food is either commercially prepared or the recipe assumes the use of a convenience ingredient. In addition, items in the FNDDS, such as dried pasta, breads with fruit or nuts, and yogurt with fruit syrup added, are broken down to more basic ingredients (e.g., flour, water, oil, and eggs) to facilitate nutrient analysis. However, CNPP assumed households purchase these foods in a ready-made form. The review also indicated that some recipes appeared to break down convenience-packaged items, such as macaroni and cheese or skillet meals in a box, into individual parts. CNPP staff modified these recipes as discussed in the next paragraph.

*Revised Recipes:* The review of recipes in the FNDDS led to a modification of 286 recipes. Modification was done either because the recipe contained ingredients indicating commercial preparation or to allow for a convenience ingredient such as a pie crust or mix. Some ingredients, such as industrial shortening, high fructose corn syrup, and hydrogenated cotton seed and soybean oil, are used only in the food-processing industry and not in home preparation, thus indicating the use of a convenience ingredient or a ready-to-heat product. In modifying the recipes, CNPP either replaced these ingredients with common household ingredients that perform the same function in the recipe or grouped several ingredients to form a commercially prepared ingredient such as a bottled sauce or a frozen pie crust. Recipes with detailed ingredients were replaced with convenience alternatives. For example, a boxed biscuit mix replaces flour, baking soda, baking powder, sugars, and table salt. Those recipes are mainly modified based on currently available recipes from Web sites, such as Betty Crocker (<http://www.bettycrocker.com>) and My Digital Kitchen (<http://mydigitalkitchen.com>).

*Automation:* Much of this step is automated for future updates of the price database. Only the new recipes in the FNDDS 2.0 will require an evaluation as to how the food will be included in the CNPP prices databases (e.g., as ready-to-eat, ready-to-heat, using commercially prepared convenience items in preparation, or using the FNDDS recipe). The SAS version 9.1 (SAS Institute) programs will be used to scan the new recipes for ingredients indicating a convenience or commercially prepared item. Any new recipe that is not flagged in this initial scan as needing modification will need to be reviewed manually to ensure that the additional convenience items are not required. In addition, to improve the recipes, further research is needed on how Americans prepare their foods.

## **2. Apply Yield Factors.**

*FNDDS and Revised Recipes:* Applying the yield factors allows calculation of the quantity of each prepared ingredient present in 100 grams of the as-consumed food. The yield factor consists of three parts: the refuse loss, the moisture loss and gain, and the overall moisture and fat loss and gain. The first two are applied to the ingredients in the recipes to calculate the quantity of the ingredient as purchased. The third is applied to all ingredients once they have been assembled. For example, the gram weight of the rice, pasta, and vegetables in most recipes is the weight of the food after it has been cooked and the nonedible portions have already been removed. Rice and pasta are generally purchased in dry format, while fresh fruits and vegetables were assumed to be purchased with the skin, seeds, and peels. Many dishes are baked, fried, or simmered once all ingredients have been prepared individually. The moisture and fat lost or gained through these processes are accounted for in the overall moisture- and fat-loss and gain factor. An example calculation for meatless lasagna with vegetables is given in table 1 and will be discussed in more detail as each individual step is described.

The *refuse-loss factor* is the percentage of the purchase weight that is not edible. Typically, this includes the nonedible peels and seeds of fruits, the tougher leaves and peels of vegetables, the shells of eggs, and the bones and skin of meats and poultry, if the skin is

purchased but removed before consumption. We used the refuse factors in Standard Reference 18 for most products. The exceptions were fish and shellfish, nuts, and canned goods. For fish and shellfish, we assumed that fish were purchased as fillets without bones and shellfish were purchased without the shell. Thus, fish and shellfish had no refuse loss. Similarly, nuts were assumed to be purchased without the shells and thus had no refuse. Canned goods are listed in both Standard Reference 16.1 and 18 as drained, but no refuse factor for the juice or water is given; the refuse factor was obtained from *Agriculture Handbook No. 102* (USDA, 1975). In the lasagna example, the peels and ends are removed from the broccoli, carrots, and onions; the nonedible ends are removed from the summer squash; the seeds and top are removed from the peppers; and the roots are removed from the spinach (table 1). The table indicates that 39 percent of the purchased broccoli weight is not considered edible.

**Table 1. Calculating the purchase amount of ingredients in meatless lasagna with vegetables**

Ingredient	Recipe FNDDS	Recipe to make 100 g	Refuse factor	Moisture-gain factor <sup>1</sup>	Purchase amount
	(g)	(g)	(%)	(%)	(g)
Cheese, mozzarella, part skim	226.8	6.50	0	0	7.39
Cheese, parmesan, grated	25	0.72	0	0	0.81
Cheese, ricotta, whole	453.6	13.01	0	0	14.78
Garlic powder	1.4	0.04	0	0	0.05
Spices, oregano, dried	1.5	0.04	0	0	0.05
Salt, table	12	0.34	0	0	0.39
Broccoli, boiled and drained	78	2.24	39	1.01	4.19
Carrots, boiled and drained	77.6	2.23	11	0.84	2.39
Mushrooms, boiled and drained	78	2.24	3	1.18	3.09
Onions, raw	80	2.29	10	0	2.90
Peppers, sweet green, boiled and drained	68	1.95	18	1.33	3.60
Spinach, raw	141.75	4.07	28	0	6.42
Squash, summer or zucchini boiled and drained	90	2.58	5	0.98	3.03
Tomatoes, red, canned, whole, regular package	720	20.65	0	0	23.46
Tomato paste, canned, with salt	170.1	4.88	0	0	5.54
Sugars, granulated	25	0.72	0	0	0.81
Macaroni, cooked, enriched	1238.3	35.51	0	0.42	16.96
<b>Total</b>	<b>3487.05</b>	<b>100.00</b>			<b>95.85</b>

<sup>1</sup>A moisture-gain factor less than 1 indicates the product will gain moisture when cooking the ingredient from the as-purchased form to the form listed in the ingredient list. A moisture-gain factor greater than 1 indicates the ingredient will lose moisture.

The moisture loss or gain is the amount of moisture a food gains or loses during the cooking process. Rice and pasta gain moisture during cooking, but many other foods lose moisture. Unlike the refuse loss, the moisture factor is not a simple percentage. We calculated the moisture factor by using a mass-balance equation from chemical engineering:

$$\text{Moisture factor} = \frac{(100 - \text{percent moisture in cooked form})}{(100 - \text{percent moisture in raw form})}$$

The information on moisture content is available from a variety of sources. Our preference was to use Standard Reference 18 or 16.1. However, information on the moisture content of raw forms was not available for all foods in the Standard Reference, so CNPP used *Agriculture Handbook No. 102* (USDA, 1975). In a few cases, CNPP staff estimated the moisture loss based on a similar food. In the lasagna example (table 1), the broccoli, mushrooms, and peppers lose moisture (from the as-purchased form), while the carrots, summer squash, and macaroni (lasagna noodles) gain moisture as these items are cooked. Note that a moisture factor greater than one indicates that the product will lose moisture between the purchased form and the form used in the recipe. Thus, the consumer will need to purchase more of the product.

Once the refuse and moisture factors were determined for all ingredients, CNPP calculated the purchase amount of the ingredient. This is the quantity of the ingredient that must be purchased to obtain 1 gram of the ingredient in the form given in the recipe. This number is calculated by using the following formula:

$$\text{Purchase-amount factor} = \text{moisture factor} * \frac{100}{(100 - \text{refuse-loss factor})}$$

The purchase-amount factors are not shown in table 1. For broccoli, the purchase-amount factor is 1.65, meaning that for every 10 grams of boiled and drained broccoli in the recipe, the consumer will need to purchase 16.5 grams of raw broccoli. However, one more step is needed to calculate the total amount of each ingredient required to make 100 grams of the food. We must estimate the change in quantity of each ingredient from cooking, frying, simmering, or otherwise heating the entire food once assembled. The FNDDS contains the overall moisture- and fat-loss and fat-gain factors for each as-consumed food that was reported. For example, many meats lose fat when grilled, whereas vegetables gain fat when fried. In the lasagna example, baking the assembled lasagna results in an overall moisture-loss factor of 1.14—meaning that about 12 percent of the moisture contained in the lasagna evaporates as it bakes. Thus, the final purchase amount is as follows:

$$\text{Purchase amount} = (\text{quantity to make 100 g}) * (\text{purchase-amount factor}) * (\text{overall moisture- and fat-loss/gain factor})$$

The lasagna example in table 1 provides a summary of these steps, which incorporate the use of refuse-loss factors as well as the moisture gain and loss. The first column shows the number of grams given in the recipe for lasagna as presented in the FNDDS. This will make nearly 3,500 grams of vegetable lasagna. The second column shows the number of grams of each ingredient required to make 100 grams of lasagna. The pasta and all the vegetables are in their cooked form in the recipe. The vegetables have had stems, peels, and seeds removed; the pasta has gained significant amounts of water. The percentage of the total purchase weight that was stems, peels, and seeds is reflected in the column labeled *refuse factor*. Percentages for refuse loss in this lasagna recipe range from zero for the non-produce items to 39 percent for the broccoli. The moisture-gain/loss factor shows the multiplier for the recipe amount to derive the purchase amount. A number greater than one indicates that the ingredient loses moisture during preparation; thus, people must purchase more than the recipe requires. A number less than one shows that the ingredient gains moisture, so people should purchase less than the recipe requires of the as-consumed form. The final column is the amount to purchase once the refuse factor, moisture-loss/gain factor, and the overall moisture- and fat-loss/gain factors for the lasagna food code have been applied.

*Commercially Prepared Foods:* The method described previously covered the foods that required preparation. CNPP assumed other foods were purchased in a ready-to-eat or ready-to-heat form. These foods could be consumed without refuse loss or the addition of other ingredients, including water, but may require heating or cooking in a microwave, toaster (toasted breads), oven, or on a stove. Many of these foods also have moisture losses, and in some cases, fat losses. Because there are no other ingredients added to them, there are no gains (moisture or fat) for individual ingredients. The FNDDS moisture- and fat-loss factors were used to calculate the quantity of the food people would need to purchase to have 100 grams of the food in the as-consumed form. In a few cases, the appropriate factor from FNDDS was missing; consequently, the factor from *Agriculture Handbook No. 102* was used.

*Automation:* The SAS programs are designed to use the newer versions of the FNDDS and Standard Reference data files, with only minimal reprogramming. The moisture-gain factor for individual ingredient calculations requires having a cooked and uncooked form for the ingredient. These forms were not available for all ingredients; so, an estimation was made for the moisture-loss/gain factors. For the next update of the CNPP Prices Database, the refuse and moisture-loss/gain factors will need to be reviewed for about 50 foods, as well as for new ingredients.

### **3. Create the Final Recipes.**

The final recipes contained the number of grams of each purchasable ingredient or ready-to-heat item required to make 100 grams of each as-consumed food code. This list of ingredients is used to calculate the final price of the food. This step is fully automated, because it is simply a matter of combining individual data files.

#### **4. Create List of Items To Purchase.**

The final step in creating recipes appropriate for calculating the cost of a food was to create a list of items that would require a price. This list consisted of purchasable ingredients from the recipes developed in the previous step, as well as the foods purchased in a ready-to-heat or ready-to-eat form. Creating a list of items to purchase is also fully automated. This list was used in the next step, matching ingredients to items in the Nielsen Homescan™ Consumer Panel data.

#### **5. Match Purchase Items To Price Data.**

As with any attempt to merge two very different data sets, the match between USDA foods and Nielsen Homescan™ data was a difficult process. As discussed earlier, the USDA food databases are developed and maintained with the primary purpose of estimating the nutrient content of as-consumed foods reported in dietary intake surveys. On the other hand, the Nielsen Homescan™ panel data are used primarily by the food industry to track the sales of products or types of products across different demographic groups for marketing purposes. The level of details about food is different between the two databases. For example, Nielsen Homescan™ panel data distinguish carefully between flavors of tomato-based pasta sauces, whereas USDA groups all these pasta sauces under marinara sauce. To the extent the NHANES respondent understands and reports the difference, USDA distinguishes between whole- and refined-grain products. However, Homescan™ data do not differentiate as clearly. The overall goal for the match was to get a good estimate of the price of an ingredient or food in the USDA database without greatly compromising the ability to use the matches to examine nutrient content in Homescan™ foods for economic analysis.

The matching process began by manually finding the set of Nielsen Homescan™ product attributes that best matched each ingredient or food from the ingredient list compiled in the previous step. As discussed in the data section, the product attributes describe the product category (fluid milk), the type (skim), flavor (chocolate), brand, and other characteristics. In some cases, up to five different product attributes were required to obtain a good match. If the product attributes did not allow a clear enough breakdown, the match process also involved scanning through the UPC description for certain key words. This scanning was particularly needed to distinguish whole-grain (or partially whole-grain) products from refined-grain products and full-fat from low-fat products. This last step was undertaken only if there were notable differences in price between the two versions of the same product. The matches to price data were done by a team of economists and nutritionists, and all matches were verified by a USDA staff member from the Food, Nutrition and Consumer Services who did not perform the original match. Differences were resolved by the entire team of matchers.

After the matches were established manually, SAS programs were used to link the purchase data set to the list of ingredients required to prepare the as-consumed foods reported in NHANES. Purchase data from all households in the portion of the Homescan™ panels provided by ERS were included in this match. Approximately 700,000 different products<sup>3</sup> were used to estimate the price of the foods reported as having been consumed.

---

<sup>3</sup>Products are distinguished from one another by their Universal Product Code (UPC) or the code assigned by Nielsen for products sold by random weight.

*Assumptions in creating the match:* Considering the degree of complexity of both the USDA dietary data and the Nielsen data, the assumptions required to create the matches were remarkably small. However, four assumptions were made in linking the two data sets. First, if products that are considered different in the USDA data could not be easily distinguished in the Homescan™ data and a preliminary check of market prices showed no difference in price, we did not assume a difference in price. Thus the following items have the same price: low-sodium canned goods and regular-sodium canned goods; tuna packed in oil and tuna packed in water; decaffeinated and regular coffee; and different flavors of yogurts, toaster pastries, cake mixes, crackers, and skillet-type boxed meals. However, the prices for the different fat content of milk are reflected in the database because the prices vary, and both food consumption and Nielsen data document the distinction.

Second, we assumed that the percent of organic foods reported purchased in the Homescan™ data was representative of the U.S. purchase habits of organic foods. This second assumption is necessary because the USDA databases do not distinguish between organic and nonorganic foods. All products that meet the description of the match and are reported as having been purchased are included in the price calculation, including those labeled organic. As reported in the next section, the price is weighted by frequency of purchase. For most products, the organic purchases represented a small percentage of the purchases and had minimal effect on the average price. However, if a significant number of participants purchase organic foods, price will be affected.

The third assumption deals with cuts of meat purchased as random-weight items; that is, the weight of the food item varies with each purchase. CNPP assumed that the random-weight packages of chicken are purchased with the skin. Thus an appropriate refuse factor is applied to reflect skin removed by the consumer. In addition, CNPP assumed a single trim or fat content for the less common cuts of meat. The Homescan™ panel data for random weight packages of chicken in 2001 and 2002 did not clearly distinguish between chicken purchased with the skin and without the skin. Similarly, the UPC description and the product attributes for these foods do not capture differences in trim or fat content for some of the less common meat cuts. More common cuts are also sold in packages with a processor's UPC. The UPC description allowed CNPP to distinguish among cuts of meat, and the price reflects differences in fat content or trims.

Fourth, since USDA Food Plans do not include alcohol, alcohol used in recipes was priced as apple juice. Prices were not calculated for alcohol consumed as a beverage. Since only a few recipes contain alcohol, this substitution was not made often.

*Automation:* This step will require more time to complete in future updates. New foods will need to be matched and verified. Any match depending on a scan of the UPC description will need to be redone because the abbreviations change every year. There are also changes to the product descriptions within the Homescan™ panel. New product modules and attributes are added and changed each year, forcing changes to the match. Finally, CNPP estimates that approximately 20 percent of existing matches should be redone with each update. This will

allow a complete overhaul of the matches every 10 years. However, the SAS programs should continue to be used to merge the two files.

*Nutrient analysis of products within Homescan:* Because all ingredients within USDA databases contain nutrient information, it would be possible to use the matches established for the prices database to bring nutrient information into the Homescan database for the products in Homescan that match to ingredients in the USDA databases. However, there are some limitations. These matches were established for the purpose of estimating prices; there are certainly food products within Homescan that do not link to any ingredients or foods within USDA databases. As was previously described, there are products, such as low-sodium canned goods and various flavors of marinara sauce, for which the nutritional information is not product specific. Conversely, some foods such as breakfast cereals and frozen entrees for which the match is done by brand name, nutrition information brought into Homescan™ data from NHANES would be correct. Common cuts of meat, poultry, and fish; milk and dairy products; most beverages; and many common produce items have very close matches in both data sets that allow for accurate nutrition information. One last word of caution: the Pyramid Equivalent Database is calculated on foods as they were consumed. Thus, pyramid equivalents can only be calculated for single-item recipes. The analysis would need to apply the purchase-amount factors for foods that have waste, such as produce or meat or items that gain or lose moisture or fat during cooking.

#### **6. Convert Purchase Quantities to Grams.**

As indicated previously, the recipes listed the quantity of an ingredient or food item in grams. However, the purchased quantity metric is one of four possibilities: ounces or pounds, fluid ounces, counts or bunches, and quart equivalents of a powdered drink mix. The most common metric is ounces and pounds and includes many random-weight products and most grocery items. The conversion to grams for items specified in number of ounces or pounds purchased follows the standard conversion of 28.35 grams per ounce. All liquid items are sold in fluid ounces, and most of the factors to convert these items to grams are found in Standard Reference data. Items such as bagels, eggs, and some produce are sold by the number purchased or the count. Standard weights for many items in purchasable form are given in Standard Reference files. In the case of produce, we assumed the weight of a medium piece—such as a medium apple. Other weights, such as the weight of a bunch of celery, were obtained from State Agriculture Extension Service Web sites. The weight of eggs was calculated by using the UPC description size and the gram weight in Standard Reference 18 for each size. The final quantity measure, quart equivalents, is used for powdered drink mixes and reflects the number of quarts the powder in the package will make. The conversion factors for these items were calculated from data found in Standard Reference files and at manufacturers' Web sites.

*Automation:* In future updates, the new Standard Reference database can easily be applied to the purchased products, with minimal re-coding. However, it is expected that some weights for new products will need to be researched.



## 7. Calculate Prices.

Previous steps have described the development of recipes that contain the number of grams of an ingredient in the as-purchased form required to make 100 grams of a food in the as-consumed form, how the ingredients were matched to price data, and how the quantities purchased were converted to grams. The final task was to calculate a national average price for each product: (1) calculate the price of the individual ingredients (or ready-to-eat and ready-to-heat items), and (2) calculate the cost of the food.

Nielsen recommends a minimum of 75 price observations for any calculation of price (Nielsen, 2005). The purchases with fewer than 75 observations included rare and expensive cheeses, raw grains, and dried herbs and spices. Although dried herbs and spices are commonly consumed, they are not purchased frequently because a single purchase may last for several months. Because CNPP required a price for all foods consumed, alternative methods of obtaining prices for these less frequently purchased items were considered: (1) survey the Washington, DC, area grocery stores, or (2) create a proxy price based on ingredients with more observations. Food prices in Washington, DC, are above the national average (Aten, 2006), making them a poor proxy for national average prices. The second option was attempted, but CNPP discovered that if the ingredients were similar enough to be reasonably represented in price by another ingredient, the number of observations of each of the two ingredients proved to be very similar, meaning the price would be represented by few observations anyway. In a few cases, the matches were revised to obtain more observations. In the end, less than 10 percent of the ingredient prices were calculated from purchase data with fewer than 75 observations. The limited purchase data indicate that these ingredients, such as caviar, raw nopales (cactus), and ethnic foods, are used in small quantities by Americans.

The price per gram of each ingredient was then multiplied by the number of grams given in the revised recipes to derive the cost of the ingredient for the recipe. The cost of 100 grams of the food as-consumed was the sum of each of these ingredient costs. As discussed previously, this number accounts for refuse loss, moisture and fat loss and gain from cooking, and other preparation changes. Because of missing ingredient prices, CNPP was not able to calculate a price for less than 5 percent of all foods. CNPP attempted to find a similar food that did have an estimated price for these foods. In some cases, the similar food was the same food with a different code and recipe. In others, the price was matched to a more general food. For example, a frozen meal of Swiss steak with gravy, vegetable, potato, and dessert was matched to the more general frozen beef dinner. The final prices database thus contains a national average price for each as-consumed food reported in NHANES 2001-02, calculated from purchases made from all households in the Nielsen Homescan™ consumer panels from 2001 and 2002.

*Automation:* Other than defining proxy prices for any new foods where a price cannot be calculated, this step uses SAS programs to complete the process. Ingredients or commercially prepared foods with few observations will also need to be reviewed to determine whether a better match can be found.

## Conclusions

The CNPP Prices Database contains nationally representative prices for foods in the NHANES 2001-02 data set. In this first release of the CNPP Prices Database, CNPP made assumptions required for the estimation of the USDA Food Plans: the prices should reflect national average prices, and all foods should be prepared at home. CNPP priced many of the most commonly consumed foods as ready-to-eat or ready-to-heat and incorporated convenience items into recipes, particularly in cases where the recipe included ingredients (e.g., industrial shortening) not available in grocery stores. The matches of ingredients to price data were done by a team of nutritionists and economists, and all matches were verified by another USDA staff member. Researchers can now compare the average cost of groups of foods to gain a better understanding of the role of food costs in relation to consumer behavior.

However, two limitations are noted. First, no household or consumer actually faces national average prices. And unlike variation in the taste, quality, or nutrient content of many foods, the prices faced are readily available every time the consumer purchases food. The consumer does not need to depend on past experience or additional information to know the price. Using data that helped to calculate the Consumer Price Index, Aten (2006) estimates that in 2003 the range of price levels for food in metro and nonmetro urban areas ranged from 0.85 in the Cincinnati, Ohio, area to 1.29 in New York City. Thus, if a grocery cart of food items costs \$85 in Cincinnati, the same basket of items will cost \$129 in New York City. Seasonal variation of fresh produce may be even greater. Thus, caution should be used when analyzing consumer choices between foods that exhibit a great deal of regional or seasonal price variation because consumers are not facing the same information as that in the CNPP Prices Database.

Second, the prices in the CNPP Prices Database reflect the cost of foods prepared at home, even if the food clearly came from a restaurant or other sources of food away from home. As discussed at the beginning of this paper, the CNPP Prices Database was originally created and will be maintained to support the calculation of the USDA Food Plans. These plans reflect the cost of food at home in order to best support the Food Stamp Program and the U.S. Department of Defense, as well to establish child support and payments to families who foster children. Although we attempted to find frozen substitutes for these foods, the price does not reflect the cost of foods consumed from away-from-home sources, such as restaurants and fast-food places.

Future updates may include further evaluation of common cooking practices for the commonly consumed foods in American households and include more convenience items in the CNPP Prices Database. The price of the food would then better reflect the way Americans prepare food at home. In the meantime, the CNPP Prices Database provides useful data for researchers wishing to add data on the cost of food to analysis of NHANES 2001-02 dietary intake data. CNPP plans to begin estimating prices for NHANES 2003-04 dietary recall data in the near future. This will also be an opportunity to refine further the automation process.

## Glossary

**As consumed:** The form the food is in after it has been prepared and cooked. The peels, skins, seeds, bones, shells, and other nonedible portions have all been removed; the food has been fried, sautéed, baked, grilled, boiled, or broiled, and if appropriate, re-chilled.

**As purchased:** The form of the food when it is purchased from a grocery store, including the nonedible parts such as seeds, skins, and bones.

**Commercially prepared:** Foods that are prepared in a food processing facility or in-store bakery or deli, such as frozen entrees and snacks, breads, canned soups, and deli salads and entrees.

**Convenience foods:** Products that save time in preparation, such as bagged salad mixes; bottled sauces for pasta, rice, and vegetables; seasoning packages; skillet meals; frozen vegetables; and mixes.

**Moisture and fat loss/gain:** Foods lose or gain moisture and fat in the cooking process. Pasta, rice, and other grains typically gain moisture. Most fruits and vegetables lose moisture while cooking. Frozen foods lose moisture from the purchase weight, as do most foods baked in the oven. Many meats lose fat while cooking; some other foods absorb the cooking fat. The loss and gains are measured in terms of the weight of the water or fat lost or gained in the preparation and cooking process.

**Moisture-loss/gain factor:** The number used in calculating the moisture loss and gain of individual ingredients. It is derived from a mass balance equation, common in chemical engineering. A number greater than one indicates that the food item loses moisture in the preparation process, and the consumer will need to purchase more of the food in the as-purchased form to prepare enough of the recipe form. A number less than one indicates that the food gains moisture during preparation, thus the consumer purchases less.

**Overall moisture- and fat-loss/gain:** The number applied to all ingredients in the recipe to estimate the amount of moisture or fat lost and gained after all ingredients are prepared and the dish is assembled. Generally this is moisture lost from baking, but some dishes are fried, resulting in a fat gain.

**Purchase amount:** The amount of food in its purchased amount, measured in grams. This includes the peels, skins, shells, and seeds. The weight does not include the package weight.

**Recipe ingredients:** Items used in the preparation of other foods. In the Food and Nutrient Database for Dietary Studies recipes (USDA, ARS, Food Surveys Research Group, 2004), the weight of many ingredients is not in the as-purchased form. For example, in lasagna, the noodles are cooked and the vegetables are washed and the nonedible portions removed.

**Ready to eat/ready to heat:** Foods that can be eaten as they are when purchased, or that only require heating. No additional ingredients or preparation are required, but a pot, pan, or microwave bowl may be required to heat. Examples of these foods include frozen foods; canned soups; some prepared produce items such as cut fruit, baby carrots, or bagged lettuce.

**Refuse loss:** The part of the purchased food that is not consumed. Refuse includes skins, peels, seeds, shells, bones, canning juices or water, and fat trimmed by the consumer.

**Refuse-loss factor:** The percentage of the purchased weight of a product that is not edible because it is a shell, seed, unconsumed skin or peel, canning juice or syrup, bone, or fat trimmed by the consumer. This is generally applied to fresh produce, meats, poultry, and canned fruits and vegetables.

**Yield factor:** The calculated factor to take a food from the as-purchased amount to the as-consumed amount. It includes both the refuse-loss factor and the moisture- and fat-loss/gain factor. The final yield factor also includes the overall moisture and fat loss and gain once the food is assembled and cooked, such as the moisture lost from baking a casserole or simmering a skillet meal.

## References

- Aten, B.H. (2006). Interarea Price Levels: An Experimental Methodology. *Monthly Labor Review*, 47-61.
- Bowman, S.A. (1997). A Methodology to Price Foods Consumed: Development of a Food Price Database. *Family Economics and Nutrition Review*, 10(1), 26-33.
- Carlson, A., Lino, M., & Fungwe, T. (2007). *The Low-Cost, Moderate-Cost, and Liberal Food Plans, 2007*. Alexandria, VA: U.S. Department of Agriculture, Center for Nutrition Policy and Promotion.
- Carlson, A., Lino, M., Juan, W., Hanson, K., & Basiotis, P.P. (2007). *Thrifty Food Plan, 2006*. Alexandria, VA: U.S. Department of Agriculture, Center for Nutrition Policy and Promotion.
- 2001-2002 National Health and Nutrition Examination Survey. (2004).
- Lino, M. (2007). *Expenditures on Children by Families, 2006*. Alexandria, VA: U.S. Department of Agriculture, Center for Nutrition Policy and Promotion.
- Mancino, L., & Newman, C. (2007). *Who Has Time To Cook? How Family Resources Influence Food Preparation*. (No. 40). Washington, DC: Economic Research Service, USDA.
- National Research Council. (2005). *Improving Data to Analyze Food and Nutrition Policies*. Washington, DC: The National Academies Press.
- Nielsen. (2005). *Nielsen Homescan™ Consumer Panel*, from <http://www.acnielsen.com/products/reports/homescan/>.
- SAS Institute. SAS/Base (Version 9.1). Cary, NC, USA: SAS Institute, Inc.
- U.S. Department of Health and Human Services & U.S. Department of Agriculture. (2005). *Dietary Guidelines for Americans 2005* (6th Ed.). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Agriculture. (1975). *Food Yields Summarized by Different Stages of Preparation Handbook (Agriculture Handbook No. 102)*. Washington, DC: Agricultural Research Service.
- U.S. Department of Agriculture, Agricultural Research Service. (2004). USDA National Nutrient Database for Standard Reference (Version Release 16.1). Beltsville, MD.
- U.S. Department of Agriculture, Agricultural Research Service. (2005). USDA National Nutrient Database for Standard Reference (Version Release 18). Beltsville, MD.
- U.S. Department of Agriculture, Agricultural Research Service, Food Surveys Research Group. (2004). FNDDS USDA Food and Nutrient Database for Dietary Studies, 1.0. Beltsville, MD.

