Purchase to Plate Suite – Overview

The Purchase to Plate Suite is a result of linking USDA’s nutrition datasets with IRI’s grocery scanner data. There are two primary benefits that result:

1. Nutrition can be estimated for common grocery items;
2. Prices can be estimated for foods found in USDA’s nutrition databases (FNDDS and NHANES).

The **Purchase to Plate Crosswalk (PP-C)** links IRI grocery scanner data with many foods found in USDA’s nutrition databases. Since many foods found in USDA’s nutrition databases are ‘prepared’ foods (i.e. cooked, skinned, etc.), the PPC also has conversion factors associated with each grocery item that converts the purchased weight to edible weight. These conversion factors allow nutrition to be more accurately estimated for these grocery items, and by consequence, prices for the USDA foods.

The **Purchase to Plate** **National Average Prices (PP-NAP)** provides prices for many foods found in USDA’s nutrition databases by using the linkages established in the PP-C. These prices are based on grocery store scanner data and consumer purchase history and linked to the USDA nutrition databases. The primary purpose of the PP-NAP is to estimate the cost of market baskets for the USDA Food Plans, including the Thrifty Food Plan. Each price series covers two years of data (i.e. 2015/2016).

The public-use file includes the PP-NAP data from 2011/2012, 2013/2014, and 2015/2016, and is further separated between retail (prices based from IRI InfoScan) and household (prices based from IRI Consumer Network). The panels provided have 5 columns:

* FoodCode – The code that links to NHANES/FNDDS identifier for the food
* Method – The number of the method used to calculate the price
* Method\_Description – The name of the method used to calculate price (see documentation)
* NHANES – indicates whether the price and recipe are found in the NHANES data (NHANES or Top90) or exclusively to the FNDDS/SR data (Extra)
* Price\_100gm – The price per 100 edible grams of the food.

Also included in the public-use files is […]

[PUBLIC USE FILES HERE]

Purchase to Plate National Average Prices – Overview

The Purchase to Plate National Average Prices (PP-NAP) provides prices for foods found in USDA’s nutrition databases and is a result of linking scanner data from IRI InfoScan with nutrition data from the Food and Nutrient Database for Dietary Studies (FNDDS), the National Nutrient Database for Standard Reference (SR), and by consequence, the dietary component of the National Health and Nutrition Examination Survey has a substantial amount of linkages. The sales data in IRI is used to estimate prices for grocery store items from IRI, then by linking IRI with NHANES, prices can be estimated for the foods reported eaten in NHANES.

**A quick note on the relation between NHANES and FNDDS:** What We Eat in America (WWEIA) is the dietary component of NHANES, which includes one or two 24-hour dietary recall interviews. The USDA then calculates the nutritional content for the foods reported and these foods are included in the FNDDS database.

The PP-NAP is especially useful for estimating the cost of market baskets and is used by the USDA to estimate its four food plans: Thrifty, Low, Moderate, and Liberal. The Thrifty Food Plan being especially consequential, as it is used as the basis for the Supplemental Nutrition Assistance Program (SNAP) maximum benefit allotments. The PP-NAP can offer an additional dimension to NHANES research by giving users the ability to estimate food expenditures.

The PP-NAP has several limitations:

* They prices are national averages and so they do not represent prices all consumers face.
* The retail price panel does not include prices for all ‘private label’ (store brands) items as some stores do not report quantities sold on this data.
* [Consumer Network random weight items placeholder]
* [Coffee example? Store bought bottle coffee vs ground, ask Andi?]

The PP-NAP can be linked to FNDDS or NHANES data by joining the Foodcode in PP-NAP with Foodcode in those two datasets. The most important columns in PP-NAP are ‘Foodcode’ and ‘Price\_100gm’, the other columns give additional information on the price or food item.

Purchase to Plate National Average Prices – Documentation

Methods Used to Produce Data

The datasets used to build PP-NAP are USDA’s What We Eat in America/National Health and Nutrition Examination Survey (WWEIA/NHANES), the USDA Food and Nutrient Database for Dietary Studies (FNDDS), the IRI InfoScan and Consumer Network datasets, and the Purchase to Plate Crosswalk (PPC). The generalized process for construction of the national average prices is as follows:

1. Select foods from WWEIA that were reported eaten more than 10 times
2. Review recipes in the FNDDS to determine if modification is needed
3. Calculate average sale prices for each item in the IRI data
4. Convert purchase weight to edible weight using PPC, then adjust price from step 1
5. Adjust average price to average price per 100 edible grams
6. Link IRI UPCs to food codes in the USDA Nutrition databases (FNDDS and SR) using the PPC
7. Average UPC price per 100g linked to each WWEIA/NHANES food code to create a final price

The resulting prices are nationally representative of foods reported consumed by NHANES/WWEIA participants. The prices of the ingredients represent all package sizes and brands, and in many cases all flavors and types of the ingredient. The IRI data are listed at the individual product level as purchased in a grocery store. The WWEIA/NHANES codes and supporting recipes in the FNDDS are more general (e.g. BBQ sauce). More information on each of these datasets is included below.

Sources of Data

**National Health and Nutrition Examination Survey/What We Eat in America (NHANES/WWEIA) ­**– NHANES is a biennial program of surveys aimed at assessing the health and nutrition of adults and children in the US. Of particular interest is the What We Eat In America (WWEIA) survey that includes one or two 24-hour dietary recall interviews. The USDA then calculates the nutritional content for the foods reported, including the vitamin and mineral content, and the food pattern equivalents (food group quantities). Participants describe the quantity of the food and this is converted to grams of food. Note that the grams of food are only the edible portion, after the skin, peels, seeds, bones, and shells are removed. The results of this survey help determine the foods found in the FNDDS database.

**Food and Nutrient Database for Dietary Studies (FNDDS)–** The FNDDS determines the nutrient content for the foods in the WWEIA survey. The recipes in FNDDS are used to estimate the cost of the food. USDA determines the average nutrient content of these foods by developing standardized recipes. Some recipes are a single ingredient (e.g. crackers, bread, fruit, milk, juice, cooked meat with no salt added), while others include multiple ingredients. The state of the food reported, cooked or raw for example, are considered when matching with FNDDS so that nutrition is more accurately estimated. As a result, some recipes were modified to serve as a basis to estimate the price.

**IRI InfoScan and Consumer Networks –** The IRI InfoScan dataset is a record of store sales, at the bar-code level, from across the country. The IRI consumer network includes consumers who record their food purchases over a period of time. The IRI Product Dictionary (PD) accompanies the IRI Consumer Network and InfoScan Datasets, which are a record of grocery and convenience store purchases for select consumers and stores. The IRI PD offers additional information at the Uniform Product Code (UPC) level, including weight, flavor, and nutrients from the nutrition facts label. This nutrient information, however, is lacking in comparison to USDA nutrition databases. The weight is the purchase weight so items like fresh fruits and vegetables may contain refuse (e.g. peels, seeds, skins) while other items such as raw meat, pasta and raw rice would gain or lose weight when the consumer cooks the food. The statistical properties of these data are available (LINK TO USING SCANNER DATA AT ERS)

Sources of Error

There are three known sources of error in the PP-NAP: quality of the matches, choice of recipe used to price the food in NHANES, and issues with the scanner data.

1. **Matching** – Although matches between the scanner data and the USDA data are done using computer-assisted methods, each match is reviewed by trained nutritionists. And for consistency, a sample of each nutritionist’s review is reviewed by a senior nutritionist. We say the error rate for the matches is less than 5 percent, but the error is close to zero. Remaining errors in the matches are most likely due to errors in the product dictionary (see scanner data)
2. **Recipe choice –** We started with the recipe developed to estimate the nutrient content of the foods reported consumed by WWEIA/NHANES participants. However, most of these recipes had to be modified and, in some cases, we chose to use a prepared product. When different recipes of two closely related products (e.g. hot chocolate made from a mix vs a mocha-latte purchased ready-to-drink), the price differences are not considered errors.
3. **Retail scanner data**:
   1. Errors in the product dictionary – In rare cases, the product information is incomplete, or the product is categorized incorrectly
   2. InfoScan. Please see (Levin et al., 2018, Muth et al., 2016) for more details.
      1. Missing stores - while InfoScan represents half of all grocery sales in the U.S. Economic Census, only about 15 percent of stores. The coverage varies by part of the country
      2. Representativeness of stores – ERS is developing store weights to account for some of these short comings
      3. Private label – Some chains do not share their private label (store brand) data with ERS at a sufficient level of detail to estimate prices.

Purchase to Plate National Average Prices – Additional Reads

**Price Construction**

For more background on how National Average Prices are constructed please see [ERS Technical Bulletin 1955 “Estimating Prices for Foods in the National Health and Nutrition Examination Survey: The Purchase to Plate Price Tool.”](https://www.ers.usda.gov/publications/pub-details/?pubid=99294)

**IRI InfoScan and Consumer Network**

For more background on the IRI databases, including considerations when using the data, please see [ERS Technical Bulletin 1942 "Understanding IRI Household-Based and Store-Based Scanner Data."](https://www.ers.usda.gov/publications/pub-details/?pubid=47636) For comparisons of IRI InfoScan and Consumer Network databases to other databases, please see [ERS Technical Bulletin 1949 “Estimating Store Scanner Data: A Comparison of the IRI InfoScan Data with Other Data Sets, 2008-12”](https://www.ers.usda.gov/publications/pub-details/?pubid=90354) and [ERS Technical Bulletin 1946 "Food-at-Home Expenditures: Comparing Commercial Household Scanner Data from IRI and Government Survey Data,"](https://www.ers.usda.gov/publications/pub-details/?pubid=85251) respectively.

**USDA Nutrition Databases**

There are two USDA Nutrition Datasets used for the National Average Prices, FNDDS and SR, and their homepages are linked below.

* [Food and Nutrient Database for Dietary Studies (FNDDS)](https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fndds/)
* [National Nutrient Database for Standard Reference (SR)](https://data.nal.usda.gov/dataset/usda-national-nutrient-database-standard-reference-legacy-release)

**USDA Purchase to Plate Crosswalk (PPC)**

For more information on linking the IRI databases to the USDA databases, including matching methodology and conversion factors, please see [ERS Technical Bulletin 1952 “Linking USDA Nutrition Databases to IRI Household-Based and Store-Based Scanner Data.”](https://www.ers.usda.gov/publications/pub-details/?pubid=92570)