Transcription of "Using the Purchase to Plate Suite to meet USDA's Mission"
Presented February 23, 2022 by Andrea Carlson, PhD and Christopher Lowe, MS

[slide 1] I am so excited that you are interested in the Purchase the Plate Suite of products. I personally think that they are amazing set of research tools. And I look forward to seeing how you will use them to conduct research that will help USDA fulfill its mission and meet our strategic goals. In order to get the most out of this presentation, you should have a basic understanding of household and retail scanner data. There is more information on the IRI scanner data, on the ERS web page using "scanner data".

[slide 2] This is not a solo project and I need to acknowledge the current team from USDA's Economic Research Service, the Center for Nutrition Policy and Promotion, and our contractor, Westat.

[slide 3] The Purchase to Plate Suite combines datasets to enhance research related to the economics of food and nutrition. It also supports USDA projects and programs such as updating the market basket for the Thrifty and other USDA food plants. The three products use proprietary data are the Purchase to Plate Crosswalk or the PPC. And this allows scanner data users to measure the healthfulness of store purchases. The Purchase to Plate Price Tool, or the PPPT, allows users to estimate custom prices for enhanced foods. And finally, the Purchase to Plate Ingredient Tool, or the PPIT, breaks enhanced foods back into ingredients based on US retail food purchases. I am going to talk about each of these in a little more detail so don't worry if this did not make sense to you.

[slide 4] To begin, let's start with the Purchase to Plate Crosswalk. This allows scanner data users to import the nutrition data from the USDA nutrition databases. Again, I'm not going to go into detail about the nutrition databases because that information is available on the USDA Agricultural Research Service website. What you need to know is that these databases will allow you to calculate the Healthy Eating Index, or HEI scores, using scanner data. Or, you can examine nutrient content beyond what is on the nutrition facts panel. For example, one database that USDA has, the food and nutrient database for dietary studies, contains nutrient information for 64 nutrients for each item in the database.

[slide 5] For those of you who may be unfamiliar with the Healthy Eating Index or perhaps need a refresher, the basic information that you need to know is that the HEI is a measure of compliance with the USDA Dietary Guidelines for Americans. It is used by economists, nutritionists, and epidemiologists and it is updated when the dietary guidelines are updated. The HEI 2015 is the most current because the factors included in the HEI did not change very much between the previous guidelines and the current guidelines. In the HEI, there are a total of 100 points, including 13 components, and that includes nine adequacy and four moderation components. And a higher score always equals better compliance. Let me warn you right now, this will trip you up. For example, there is an added sugar component. A higher added sugars score means that the person or the basket of food has less added sugar than a basket with a lower

added sugars score. If you need more information it is available on the National Cancer Institute website, HEI website.

[slide 6] For those of you who think visually like I do, here are the steps to calculate the Healthy Eating Index. We start with the transaction data in the scanner data that links to a product dictionary, which joins with the crosswalk, which then joins to the USDA nutrition databases. And these feeds into the Healthy Eating Index program. Now, if this did not make sense to you, there is no need to panic

[slide 7] because we have created a user guide to assist you in using the Purchase to Plate Crosswalk. Included in the user guide are sample programs that illustrate how to calculate the HEI scores. We strongly encourage you to review the user's guide and use the sample code as a basis for your analysis. That way you use it correctly.

[slide 8] Now that you know what the Purchase to Plate Crosswalk does, I'd like to talk about the coverage rates. In this chart we measured coverage as the percent of total sales in InfoScan or the percent of total purchases by households in the consumer panel over three of the four Crosswalks created to date. Not shown in this slide, is the coverage for the 2017 18 TBC. But it's very similar to the 2015-16 one. For the first two Crosswalks, the one designed for the 2013 and 2015 data, the unmatched items tend to be items that did not have a counterpart in the USDA database. For the 2015-16 and the 2017-18 PPCs, we had to change our methods to meet the extreme tight timeline on the update of the Thrifty Food Plan. In addition to that, the underlying FNDDS underwent a significant change and there were more IRI products that did not match to an FNDDS item. As you can see, all of them have good coverage rates for most analysis when it's applied to the years it's designed for. However, you notice that the 2013 PPC is a little bit low for the consumer network. That is because that PPC was not originally designed for the consumer network.

[slide 9] When using the PPC for years other than the one it is designed for, it is very important that you check the coverage rate across the IRI data. For example, shown in red, is the coverage rate when the 2015 PPC is matched to the 2015 InfoScan. However, if that same crosswalk is applied to the 2008 IRI InfoScan, you can see the coverage rate drops for some categories, such as beverages, frozen foods, and refrigerated items, but not so much for meat and produce. This coverage rate is shown in blue. Even moving forward just one year to the 2016 IRI data, shown in yellow, the coverage rate drops off unevenly. These differences in coverage rates will have an impact on the HEI score. Because beverages and frozen foods have a higher percentage of items that are less healthy than produce and meat. Refrigerated items include dairy and eggs, but it also includes refrigerated prepared foods, some of which are high in saturated fat or added sugars. This all means that the estimated HEI score of retail food purchases will be higher than the true HEI. In addition, there are substantial changes in the 2019 and 2020 IRI produce product dictionaries. So if you plan to use the 2017 and 18 PPC with the 2019 and 2020 data, this may result in a very low coverage rate for produce, which will have a huge impact on your HEI score.

[slide 10] Moving on to the second tool, the Purchase to Plate Price Tool, allows researchers to use the scanner data to estimate prices, but for the foods reported consumed by participants in the What We Eat in America Study, which is the dietary component of the National Health and Nutrition Examination Survey. And you can estimate these prices using a subset of the IRI data. The purchase, the Purchase to Plate Price Tool is based on the recipes in the FNDDS and the prices themselves are drawn from the scanner data. Just a note that ERS has a forthcoming data product, the purchase to plate national average prices for NHANES (PP-NAP) which will publicly provide national average prices for NHANES foods.

[slide 11] In order to use the price tool with the scanner data that it is important to understand NHANES and What We Eat in America. NHANES or the National Health and Nutrition Examination Survey is a nationally representative survey of about 5000 individuals each year. It is released in two-year cycles, except for the year 2019 and 2020 and it includes a very detailed medical exam. It is used by thousands of researchers every year for a wide variety of studies. What We Eat in America is the dietary component of NHANES, and it includes two 24-hour dietary recalls. People just say what they ate. The amounts reported are as they are eaten, what's on the plate. The FNDDS is used to calculate the nutrient content of the foods reported by participants in What We Eat in America. And we can use it to estimate prices.

[slide 12] We started with the recipes that the FNDDS uses to calculate nutrient content. Here are two sample recipes. The first, vegetables, not specified as to type, cooked, fat not added in the cooking. This recipe is straight forward, and it can be followed using grocery store items, basically mixed frozen vegetables, and salt. But we do need to standardize it. So it makes 100 grams of the What We Eat in America food, the mixed vegetables. The second one, cooked carrots, contains a food that can be purchased, cooked carrots, but most people likely purchase their carrots as fresh, frozen, or canned. And the fresh carrots are purchased in a wide variety of formats with and without refuse. The FNDDS recipes are designed to calculate the nutrient content of the food. As a result, they contain standardized items such as cooked carrots for consistency across all recipes that includes carrots that will get boiled at some point. In order to calculate prices, we had to modify most of the recipes to represent grocery store items.

[slide 13] To make these modifications, we made some key assumptions and guiding principles. First, all the ingredients used had to be purchasable in stores. We couldn't use these kind of part way cooked foods. Second, we assumed that commonly used convenient foods, such as bottled salsas, prepared ingredients, and frozen entrees are used. Now there have been some recent changes in FNDDS, which creates a statistical representation of a nutrient content of food using a wide variety of preparation methods. This will make implementing this assumption about convenience food easier for some recipes. But the changes also mean that fewer IRI items will have a valid match in the PPC. And finally, we assume that almost all foods are prepared in 30 minutes or less. This time limit is based on ERS research using time use studies. However, the estimate for the preparation time is based on either the number of ingredients--making the assumption if there are a lot of ingredients, then it might take longer to prepare-- and some

professional judgment. Again, this professional judgment is done by nutritionists and food scientists whose job it is to just plain know this stuff.

[slide 14] While the national average prices using all of InfoScan data is a forthcoming, publicly available ERS data product, the tool is designed to allow researchers to use a subset of InfoScan to calculate the prices. For example, you could break it down by the type of store, you could use the geographic region. And to help that, ERS has new store weights that will allow you to break the RMA, the retail market area, stores back down into individual stores. You can calculate your prices that way based on geography. You could also break it down between organic and non-organic food. You could break at private label versus brand. Anything else you can think of, if you can make a subset of the info scanned data, you can use that subset to calculate these prices. However, there are some considerations that you need to keep in mind. The first is that the tool will not create a price if the price of an ingredient is missing. So, if you're thinking back to that mixed vegetable, if you don't have a price for frozen vegetables, the tool will not create a price for it that cooked not further specified vegetable. In addition, the researcher will be responsible for ensuring there are an adequate number of observations for each ingredient that goes into the price. Given that, I still really look forward to seeing the policy-oriented research you develop using this tool.

[slide 15] Our final tool is the Purchase to Plate Ingredient Tool. This is the opposite of the Purchase to Plate Price tool. We start with What We Eat in America and NHANES and work backwards to items you can purchase in a grocery store.

[slide 16] For example, if a participant and What We Eat in America reports that they had rice, but they didn't remember what name the rice that they ate, the price of rice would be based on all of these types of rice. Note that the ingredient tool breaks the rice down by purchase form, such as refrigerated, not ready to serve, shelf stable ready to serve, or frozen not ready to serve. The amounts to purchase are based on the amount sold of each type of rice in InfoScan. If you add all these amounts together, you will get the number of grams of raw rice to purchase to make 100 grams of cooked rice. Keep in mind that this slide does not include the water, which is the heaviest component of cooked rice. As you can see, the white rice is the most popular form of rice, with brown rice being a distant second.

[slide 17] In addition to the ways you could create that sub-sample of the InfoScan data to calculate your prices, you can then further sub-divide how you're going to break down the ingredient grouping. The sample code that we include, breaks it down by the IRI form, which includes shelf-stable, refrigerated, frozen, ready to serve or drink, ready to heat. with refuse-- with seeds, bones, shells, what have you-- or with no refuse, so the food is as you eat it. You could also use the new ERS food purchase groups. These are a new ERS value added product for the IRI data. These groups categorize purchased foods by ingredients, the nutrient content, convenience, and store aisle. For those of you familiar with FoodAPS, they're very similar to the FoodAPS groups. And as I said, we do have sample code available.

[slide 18] So to recap, the Purchase to Plate Suite, which is shown in green, has four components. The purchased to Plate Crosswalk goes between the scanner data and the FNDDS, which then can be linked to another USDA database and calculate the Healthy Eating Index. While the Purchase to Plate Price and Ingredient Tools link goes from the scanner data to NHANES or back the other way for the ingredient tool. And then the Purchase to Plate National Average Prices is a side product of that--the national average prices we are able to make publicly available.

[slide 19] This is a list of the years for which we have available data. Note that in the first two cycles, the 2011-12 and the 2013-14, these do not match to the same years as the IRI data. In the beginning, we match the most current FNDDS to the most current years of the IRI data. Once we realized that this was an ongoing project, we switched to align the years. And thus the 2015-16 FNDDS links to the 2015-16 IRI data. Please note that the 2020 data, which everybody is looking forward to because of COVID, will not be available until at least the fall of 2023. We first must wait for the new FNDDS to be released before we can even start the process of linking the data, creating recipes, and finalizing the sample code.

[slide 20] I also want to recap the limitations of the Purchase to Plate Suite. For the Purchase to Plate Crosswalk the match rate varies by supermarket category. You must pay attention to this. And it is not fully tested with years other than the creation year. So, make sure you check that coverage rate. For the Purchase to Plate Price Tool and Ingredient Tool, the price differences between two items might be based on the price calculation method rather than an actual change. For example, many of the coffee drinks are priced as if you bought a bottle of coffee, coffee mocha, and others assume that you brewed coffee at home. Again, these have not been fully tested with subdivisions of the IRI data so you will need to verify that it worked. And it generates average prices, which may not be completely appropriate for most demand models, since it will generate a situation where everybody is facing the same price. There are known limitations to the scanner data that apply directly to the Purchase to Plate data. The retail data represents 15 percent of stores and half of all retail sales. The purchases represent most large chains. So, any kind of breakdown you see may not represent what is available at smaller and independent stores. And finally, not all stores provide data to IRI or allow USDA to access it. And when using the household data, with the PPC, we have documented underreporting by participants. And we are not able to include the random weight data because we don't have the quantity purchased.

[slide 21] And I thank you very much for your attention. I hope that this has given you enough introduction so that you will be able to help USDA fulfill its mission and conduct policy-oriented research. Thank you.