Food and Health Research Using the Consumer Food Data System

Questions & Answers from Office Hours with ERS and Census Data Experts (on April 18, 2022)

This document lists the questions that were asked during the April 18th office hours and the associated answers provided by ERS and Census data experts. If you have additional follow-up questions about specific data assets, please email the data expert directly.

Food Access Research Atlas, Food Environment Atlas, and Food Availability Data System

For the Food Access Atlas, Food Environment Atlas, and Food Availability datasets: How current are these datasets? Do they show how behavior changed during COVID?

> For the Food Access Research Atlas, the latest data available is 2019 data. The Food Environment Atlas was last updated in 2020. These data come from a variety of sources and cover varying years. These two Atlases do not show how behavior has changed during COVID.
>
> Our Food Availability (FA) and Loss-Adjusted Food Availability (LAFA) Data Series are current through 2019. We are in the process of updating the data through 2020 or 2021, depending on the commodity of interest. There is often a lag between the time that data on specific commodities are released on the ERS website by commodity analysts that provide data for the FA and LAFA series and when they are posted on the FA website. For example, while the current FA data for meat and poultry are updated to 2019, more recent data, for 2020 and 2021, can be found on the following ERS web page:
>
> https://www.ers.usda.gov/data-products/livestock-and-meat-domestic-data/

Is there any way to download the Food Access Research Atlas data, or are they only in map form?

> Yes, there is a way. In addition to the maps online, all data included in the Food Access Research Atlas are aggregated into an Excel spreadsheet for easy download.
>
> All publicly available data on the ERS website are machine readable and open to download.

Flexible Consumer Behavior Survey (FCBS) on the National Health and Nutrition Examination Survey (NHANES)

For the FCBS in NHANES dataset: In the Call for Proposals, it states that granular geographic data is not supported by this funding opportunity. If a researcher decides to use the FCBS, what restricted data variables will be supported?

> There are no geographic variables in the public-use NHANES, including the public-use FCBS data. For this grant opportunity, proposals must use public-use FCBS data for national-level analysis. Therefore, proposals will need to use other restricted-use CFDS data to meet the restricted-use requirement of this grant opportunity.

If a researcher wants to analyze FCBS data but wanted to use restricted demographic data (i.e., disaggregated race/ethnicity variables), would this meet the restricted data use component?

For this grant opportunity, only the public-use version of the FCBS can be used. Note that a wide variety of demographic variables are publicly available. However, if there is an interest in the demographics of subpopulations—for instance, country of birth in Los Angeles County, CA—this restricted-access variable will not be available for this grant opportunity.

For this grant opportunity, at least one restricted-use CFDS data asset listed in the Call for Proposals must be used.

Are the NHANES restricted-use data accessible through the ADRF? Is there any way to link NHANES data to the IRI data beyond at the food item level?

Restricted-use NHANES are only available in NCHS Restricted Data Centers (RDCs). The individuals in NHANES cannot be linked to any other datasets, including the IRI data.

The FCBS in NHANES is specifically listed as one of the public-use dataset options. Can other parts of NHANES such as 24-hour dietary recall data be used in combination with the FCBS?

All public-use NHANES data are linkable with the person-level identifier (SEQN). If linked to the FCBS data by SEQN, the 24-hour dietary recall data, which are publicly available, may be used.

For the FCBS in NHANES, is it correct that no state/county-level data identifiers exist?

That is correct – there are no identifiers for counties, states, or census regions. For this grant opportunity, only public-use FCBS data can be used for a national-level analysis.

Is there a way to determine which food datasets have been linked to NHANES/FNDDS food codes (for example, for assessment of nutrient profiles)?

For this grant opportunity, we are not providing NHANES data beyond the public-use FCBS component. But FNDDS has been used to be linked to IRI and FoodAPS. The Purchase to Plate has made that linkage.

Can you describe what is available in restricted NHANES data compared to the public-use version?

In the restricted-use version of NHANES, you can access a variety of interview- and geography-related data that do not appear in the publicly available NHANES files. For instance, interview data and granular geographic data (latitude/longitude of residences) are accessible in NCHS Restricted Data Centers (RDCs) with an approved RDC proposal. For this grant opportunity, the restricted-use FCBS and broader NHANES files are not available.

Are data for subjects in restricted-use datasets linked to data in unrestricted datasets such as FCBS? Are the same subjects/households included in all datasets?

FCBS participants in the public- and restricted-use versions are identical. However, any restricted-used information on FCBS participants are accessible only in NCHS Restricted Data Centers (RDCs) with an approved RDC proposal. The restricted-use version of the FCBS in not available for this grant opportunity.

If a researcher wants to use the racial/ethnic subgroup data in NHANES (for specific subgroups such as Chinese American), would this count as restricted data use that is required by this grant opportunity?

If Chinese American is part of the restricted access data in NHANES, then a researcher would have to use the restricted access data. However, for this grant opportunity, you can use public access data with both the Flexible Consumer Behavior Survey (FCBS) and the Purchase to Plate National Average Prices for NHANES (PP-NAP).

There is another grant opportunity focused on the Purchase to Plate suite offered through the Gilford Center. That grant requires accessing the IRI data through ADRF and thus the PP-NAP is not included. The same webinar is posted for both this grant opportunity and the Gilford Center grant so we understand the confusion.

Eating and Health Module on the American Time Use Survey (ATUS)

Is there any way to link ATUS data to any of the restricted-use data sets (e.g., SNAP administrative data and/or FoodAPS)?

There is no direct link from ATUS to FoodAPS or the SNAP administrative data. However, you can use a matching procedure (like propensity score matching) using different sociodemographic characteristics. Both ATUS and FoodAPS are meant to be representative of the U.S. population.

Will other ATUS modules also be available, or just the Eating and Health module?

The ATUS and the Eating and Health module data are available for everybody at the BLS website. https://www.bls.gov/tus/

Purchase to Plate Crosswalk (PPC), Purchase to Plate Price Tool (PPPT), and Purchase to Plate Ingredient Tool (PPIT)

For the Purchase to Plate Crosswalk data, is the food group variable in the PPC or PPPT dataset?

You can use the PPC to import the food group equivalents (the amount of the food group in each FNDDS code) from the Food Pattern Equivalent Database (FPED). The FPED is included in the ADRF, and we provide sample code to both calculate the food group amounts and the Healthy Eating Index (HEI) scores.

In the Purchase to Plate tools, is it correct that the as purchased price includes the cost of inedible parts, and as eaten is a lower price that does not include the cost of the inedible parts?

The PPPT will create prices in the as eaten form. If the product has refuse, the quantity purchased measured in grams will be more than the edible gram, but the cost does not change. The price per gram changes but the overall cost of eating say 100 grams of chicken is the same whether you have a purchase price or an edible price. If it's 100 grams, it would be the price per 100 edible grams--without the skin and the bones, for the boneless and skinless chicken code. It is a difficult concept to grasp. If a researcher is estimating food costs for foods reported ingested by NHANES participants, then the PPPT will give the correct price and weight to align with NHANES.

What is the difference between the PPPT and PP-NAP? Where can PP-NAP be accessed if it is publicly available? If there is interest in analyzing the PPPT with NHANES, would this qualify as restricted data use even though PP-NAP is publicly available?

The Purchase to Plate Price Tool (PPPT) is a set of SAS scripts that builds the Purchase to Plate – National Average Prices for NHANES dataset (PP-NAP). The PPPT uses the Purchase to Plate Crosswalk (PPC) to join IRI scanner data with USDA's food codes. USDA's food codes come from the Food and Nutrient Database for Dietary Studies (FNDDS). These USDA food codes are also used in the NHANES What We Eat in America Survey. The PPPT then takes the sales and package weight data from IRI to calculate the price per 100 edible grams for each linked USDA food code.

The above results in the PP-NAP dataset that is a list of USDA food codes each with their respective price per 100 edible grams.

The PP-NAP can currently be accessed through the NORC Data Enclave or the ADRF Data Enclave. It is currently in the process of being publicly released to the ERS website.

A researcher can use the soon to be publicly available PP-NAP with the public facing NHANES. The PPPT within the ADRF can be used to calculate prices using a subset of the IRI data sets.

Is the Purchase to Plate Crosswalk (PPC) a database? Does it include restaurant food?

The Purchase to Plate Crosswalk is essentially a database. There are scripts associated with it so that the information in the table is properly used. It only links to the IRI scanner data, both the retail and the household data. So, it will not cover restaurant food.

Are the ERS value added food groups to merge with the IRI data in the PPC or PPPT dataset?

Yes, the ERS food purchase groups are linked to the IRI data and therefore they can be used with the PPC. And they can be used with the Purchase to Plate Ingredient Tool (PPIT) to break down purchase amounts by the ERS Food Purchase Groups (EFPGs) rather than the grouping provided.

NielsenIQ TDLinks and IRI data assets (IRI Infoscan, IRI Consumer Network, IRI Weekly Retail and Monthly Household COVID-19 Response Data, IRI Weights)

Can you say more about the contents of TDLinx? Is there a list of variables online? Can it be matched to Nielsen data if a researcher already has their own access?

There is no publicly available data dictionary for TDLinx. You must have a project in place under a third-party agreement to access this. We do have a sample table of variables that is included on the TDLinx presentation for this grant (see the additional resources listed on this CFDS website: https://www.mathematica.org/projects/consumer-food-data-system-grant-competition). Each store in TDLinx has a TDLinx code which is a numeric unique identifier. We assume if a researcher has Nielsen, because TDLinx is a Nielsen product, that technically speaking it should be feasible to link the two.

Are there any suggestion on whether scanner or household survey data like NHANES would be more accurate to capture a household's diet?

These data sets will provide different information, so specific research question will determine which is appropriate. NHANES is at the individual level, and the data do not necessarily include all

members of the household. You might have to go to the restricted access to see how many members of the household are included (though, you would need to confirm this). The IRI Consumer Network–the household panel data–provides what the household purchased at retail stores. It will not include food away from home. However, the household panel provides data for at least a year for static panel members, versus the two days in NHANES.

How clean is the data? Could you estimate the amount of time typically needed to make the IRI and Nielsen data set ready for analysis?

This is a very hard question to answer, depending on what kind of analysis is desired. It's relatively clean, but there are a lot of details and it's a very large dataset. So, the time that is required to do the processing would really depend on the research question to be answered and the type of analysis.

Commercial data are obviously very rich datasets. The assets provided do not include Nielsen item-level sales or purchase data. Sales and purchase data at the item level are from IRI. For Nielsen only, TDLinx data are offered, which are grocery store location and characteristics for food at home. We do a lot of work on cleaning the data and making it available for analysis. The data are organized and cleaned and documented. We even look at some statistical properties of these data which are available on our website. We are not just taking delivery and putting it on ADRF for use. It's gone through our data quality checks.

Are the codebooks for the IRI Consumer Network, IRI Infoscan, and Nielsen Store data available publicly?

Everything that documents what's in the data is going to be available on our website. If you go to the IRI scanner data, there's going to be a link to a technical bulletin entitled Understanding IRI household-based and school-based scanner data. So, it's a little old, it only covers through 2012. But it provides a detailed summary of the data that we have and the types of variables that are available.

Will there be new rounds of the IRI Consumer Network, IRI Infoscan, and Nielsen Store data available during the period of performance?

In terms of availability, right now we have data available through 2020. We expect any day now to get from IRI full annual delivery of the 2021 data. So, we expect that the 2021 data will be available for users by this fall. But researchers should not count on their availability for this grant opportunity.

Do the SNAP online purchases have geographic identifiers?

Assuming this question is about scanner data, it is not likely that this can be singled out from the data that we have right now. The Consumer Network data does have an identifier for a SNAP participant (self-reported). IRI (or really the National Consumer Panel, which is a collaboration between IRI and Nielsen) once a year will send out a survey asking, do you participate in any of these government programs? And it's an option. Participants can choose not to respond to it, and most of them don't respond to it—but some do. In addition, there's the method of payment which will indicate if it's SNAP.

Now for all the purchases, there's a code for the retailer that the purchase took place at. That code is basically the banner name of the retailer (e.g., Safeway, Walmart), but it wouldn't indicate which Safeway or Walmart. And some of those codes do correspond to retailers that have .com in their name. The survey is not necessarily set up to be representative of online purchases. But certainly,

those purchases do show up. But if there's a different type of data available on SNAP purchases, that's not standard about the data.

For IRI Consumer Network versus FoodAPS data: It was shown somewhere that SNAP/WIC participation was collected in IRI beginning in 2017. Given that only 2 years of program data will be available for this grant (2017, 2018), how would the total sample size compare to the 4,800+ in FoodAPS?

The Consumer Network data goes through 2020, but the Purchase to Plate Crosswalk only goes to 2018.

We are conducting some research with 2013 Consumer Network data looking at poor versus non-poor. One could use our code to determine SNAP or WIC income eligible. The script to estimate whether the household income falls above or below a given Poverty Income Ratio will be available in an AAEA selected paper for this summer's meeting. Here are a few select sample sizes: < =100% of poverty threshold is about 6% of the households in the static panel, <= 130 is 11%, and <= 200 is 28%. These numbers are from the 2013 data.

Can we identify in the IRI Consumer Network and Infoscan data how each transaction was paid for (currency used)?

For the Consumer Network data only, there is a variable that indicates the method of purchase (like cash, credit, debit check, gift card, etc.) and use of certain food assistance programs, like SNAP or WIC benefits. There is a payment method for Consumer Network only. For Infoscan, sales were just aggregated into one stop, distinguished by method of payment.

We only have method of payment for households starting 2017 in IRI data.

Also, to follow up on 2017 being the first year for SNAP or WIC participation, are you saying that the method of payment does not include EBT or WIC prior to 2017?

SNAP/WIC indicators and method of payment begins 2017.

Is there a document that shows the sample sizes for WIC households and SNAP households in IRI Consumer Network?

Among households choosing to respond to the SNAP/WIC survey, roughly 1% self-identify as WIC and 10% as SNAP.

For the IRI Infoscan data, can you share any information about the relative sample sizes across store types?

Please review the "Understanding IRI Data" technical bulletin by ERS.

Does the IRI data have information about SNAP/WIC participation for households?

Yes, they do. There are two components to IRI. There's one component that is the Consumer Network. It's a household panel. And then the other one is on the retail sales. Both are at the item level. For households, as mentioned before, for the subset of households, you have SNAP and WIC participation identifiers.

Is it possible to merge/link the different data sets (e.g. the Nielsen data with the IRI Consumer Network and IRI Infoscan data)?

Yes, Nielsen's TDLinx codes are also included as a variable in IRI's Infoscan.

At what geographic level is geographic data available for households in the IRI Consumer Network and the Nielson store data?

Household data have geographic information in the household down to the block level, and up to the census, region, and state. And just to clarify that the unit of measurement is at the household level, so it is individual households reporting purchases.

TDLinx has latitude and longitude as well as metropolitan statistical areas (MSA) information in there as well. TDLinx has a set of codes called TDLinx codes, which are driven from latitude and longitude and attached to the stores. Those are available in IRI. And because TDLinx is owned by Nielsen, they should be able to link it up to Nielsen data.

What geographic disaggregation are possible with the IRI data? Is it possible to analyze product mix or prices at the census tract level?

I think you can have prices at the store level. For a number of different retailers, there are prices available at the store level. Whether you want to make a nationally or a representative estimate at the census tract level is going to depend on the sample and depend on what you do with the data. But about half of all the sales data that we do have is going to show up at the store level. The other half is going to be at the retail marketing level. And the retail marketing is going to vary by the retailer and by geography and year. For example, all the Walmart sales in the Baltimore/DC area are aggregated at that level for a particular product. So, it would be hard to translate that from that level to a lower geographic level. And then the households, I believe you have responding households, geographic information down to the block level.

Also, we'll have the monthly food at home price database, which will provide consumer price indices for a variety of products in the ERS consumer group categorizations. And this will be available at the metropolitan area, the metropolitan level for the cities considered by the Census. And then we'll also be looking at the four census regions as well.

We have created store level weights for the store. So those will allow you either to generate national level estimates, or you can also generate estimates that are for each of the 10 major metropolitan areas and the four census regions. And the reason why they were defined that way is because to get an adequate representation of stores within each of those areas so that we could develop weights, and you can generate estimates from those. So those weights will be available. And within the weighting dataset, it actually would allow you to disaggregate the stores in the retail marketing areas into their individual estimates of sales. The dataset that comes with that serves two purposes. And the weights are currently available for 2012 through 2018. We're hoping by the end of this year, we will also have weights for 2019 and 2020. And then using those weights, we also applied those for the monthly food at home price database. So that provides estimates of average prices in six different price indices depending on your need for all of the ERS food purchase groups, and also at those 10 major metropolitan areas at the four census region level.

Currently, those data have been created for 2016 through 2018. We're in the process of adding 2019 and 2020 data to that as well. So, there will be five years, and it's on a monthly basis (so it's a relatively high frequency).

For developing estimates at the census tract level, it may be a little challenging with Infoscan data, because that presumes that you have enough stores at that level to generate an estimate. IRI is not a complete representation of all retailers in the United States. It tends to focus mostly on chain stores, so it does miss a lot of the smaller, more local types of retailers. Depending on how that data are being used, it might be difficult to work with very narrow geographic areas, but certainly possible to kind of redefine your geographic areas with the information on where the stores are. But that's really digging into a lot of the detail of the Infoscan data.

In the IRI Consumer Network data, what is the "subset" of households that have health and prescription data available (how many/how are they sampled)?

I'd estimate it's about 15,000 households that would have the health information and prescription information (it's self-reported). It's a completely opt-in survey, but it's only from people who are in the so-called static panel. Those are the folks that provide enough food purchases that IRI constantly says this is representative of their food at home or purchase and consumption. It's going to be sampled from that—so about 45% of the entire static panel.

In the IRI Consumer Network data, are household demographic variables measured or projected?

Measured.

How far into 2020 is the IRI Consumer Network data available?

Until the very last day of 2020.

In the IRI Weekly Retail and Monthly Household COVID-19 Response Data, what does "projected" mean in terms of COVID household data?

These are basically projected relative to a census baseline. The households that are responding to the National Consumer Panel, which is what makes up Consumer Network Panel, that data is used to make nationally representative estimates where the weights use the demographics and the weights are developed such that they will be doing all census targets.

Are data dictionaries available for TDLink, IRI Consumer Network, and IRI Infoscan data?

There is no publicly available data dictionary for TDLinx. You must have a project in place under a third-party agreement to access this.

In the IRI Consumer Network or the IRI Weekly Retail and Monthly Household COVID 19 Response Data: Is it possible to identify SNAP and non-SNAP households?

There is an identifier for SNAP households in the IRI Consumer Network. Again, it's self-reported. So all the words of caution that typically go along with something like that. And the IRI weekly retail and monthly household data as we get it; no, we don't have that.

Regarding IRI data: could you explain how purchases of variable weight foods are recorded?

It's complicated and it varies across time. Pre-2019 for the data that we have, they are recorded such that they're generally going to be in pounds. But there will be some exceptions where things will be counts. And there is some documentation available on the Enclaves that says which is which. But in general, everything is going to be recorded as pounds. 2019 and 2020 data, you would have to look

into the product dictionary, and there would be a volume equivalent description for that particular product, which you could use to translate the quantity sold to weight or to pounds, for example.

Also, would the trip level record contain a total amount paid, plus a SNAP flag, but no breakdown of the portion of the sale paid by SNAP vs. the household?

So a trip as we think about it is that you're going to the store and you're buying things. Some things you buy with SNAP, and some things you don't buy with SNAP. But in the Consumer Network data, that would be two trips, even though they're happening at the same time. So you would have to see that it would be the same date and same store. Trip one, maybe they pay with cash, and the method of payment would say cash. Then the second trip would be SNAP, and that's how you would separate SNAP versus non-SNAP payments for that example as we would think about a shopping trip. Now, this is where it gets complicated, because that's going to depend on the household to report that as such when they're responding as a national consumer panelist. And it might not be the case that they always do that. So sometimes you might see somebody who's identifying as a SNAP participant but never record a SNAP purchase.

IRI data: At a state or county level, could the IRI consumer data and the IRI COVID response data be linked? Such as to show pre/post COVID household purchasing changes?

I suppose it could be linked in a way. I guess the question is linked with respect to what? If you have for example, with the state that individual household resides in and makes two purchases in, and then you could match on state from the COVID response data some variables. But it's not immediately obvious to me what you would do with that. But you could certainly work in that way. Linking on like the types of products is potentially a possibility. I think the product classification between the two variables within the dataset is similar. It would take a little bit of work, I think, to make sure it's seamless.

Can IRI Consumer data be used to identify WIC participants? Also, would the income "ranges" and "categories" allow one to identify SNAP-eligible (<130% of FPL) and WIC-eligible (<185% of FPL) households?

The script to estimate whether the household falls above or below a given PIR will be available in an AAEA selected paper for this summer's meeting. Here are a few select sample sizes: < =100% of poverty threshold is about 6% of the households in the static panel, <= 130 is 11%, and <= 200 is 28%. These numbers are from the 2013 data.

IRI Consumer vs. FoodAPS: This is a general question, but what are the key advantages/ disadvantages of IRI Consumer vs. FoodAPS since they collect similar information?

FoodAPS captures data on all household food acquisitions, not just grocery retail acquisitions. That is, FoodAPS also includes food-away-from-home foods and foods acquired for free, such as from food pantries and schools. Thus, FoodAPS provides a complete picture of a household's food acquisitions, but for a one-week period. In contrast, the IRI Consumer Network provides data on household grocery retail acquisitions for one-year period. Also note that FoodAPS data were collected in 2012, whereas IRI Consumer Network Data are available for 2008 to 2020.

Another advantage of FoodAPS is that SNAP households and low-income households were specifically over-sampled to allow for research on the food acquisition choices of households that are often underrepresented in other surveys, like IRI, for example. Moreover, SNAP households that were

surveyed in FoodAPS were administratively verified. That is, households that were surveyed were linked to SNAP program administrative records to confirm that they were in fact participating in SNAP. With access to administrative records, we can additionally verify when they received SNAP program benefits in relation to their survey week.

Lastly, FoodAPS includes a geography component and links to USDA nutrition databases. These data provide comprehensive information on food access, the local food environment, and diet quality.

Can IRI Consumer data be used to calculate a Healthy Eating Index?

Yes, that is what the Purchase to Plate Crosswalk does. You can use it for both Infoscan and the Consumer Network data. However, it only goes through 2018 at this point. And during this funding period, we will not have 2019 and 2020 available.

What's the smaller geographic unit available in FoodAPS and in IRI?

For FoodAPS, we have very detailed geographic information on the households. We have geocodes for them—so as fine as you can get, as well as the places that they visit.

For IRI for household, you would know their census block. With the stores that they're purchasing, you wouldn't know any geographic information about the stores themselves. In the sales data, some data would be at the store level, and you'd have the exact address of the store. Then the other half of the sales data is going to be at the RMA or retail marketing level. That's typically going to be like a region that may span multiple states. But you would know all of the individual stores that make up the automatic. And for all those stores, you would have the address.

IRI Consumer Network. Muth et al (2016) note that households record total price paid but not quantities of random-weight foods, analysts must conduct additional calculations to develop estimates of the quantities. They mention One such approach is to calculate average price per pound (or other units) for each random-weight food from the InfoScan data and then divide the total price paid in the Consumer Network data by the calculated average price to obtain an estimate of the total pounds (or other units).

Since this report was published (2016), has this approach been validated? Or is there another approach you would recommend to estimate the quantities purchased?

This is a hard question. So yes, just to confirm for random weight foods, essentially loose weight perishable food products, households just record the total expenditure for that particular category. And the categories are broad—for example, just bananas or chicken thighs—no distinction of bone-in chicken thighs versus skin-on boneless chicken. So they would just mark the quantity as one. It's not something to be interpreted, and then record the total expenditure. To recover the quantities, I guess you could do what is suggested in Muth et al 2016.

This work was started by a former ERS staff person who was a very good programmer but unfortunately left. So there’s a lot of work to tease it out. The idea was to come up with prices from Infoscan and then go to Consumer Network and generate the quantities there. So conceivably, that can be done by somebody, but it hasn't been done here. In theory, it should work but there's going to be a lot of caveats to it. And you might want to get yourself somebody who's good with SQL.

Re: IRI Consumer Data: Is SNAP/WIC participation measured "currently" or asked about the last 12 months?

Reference point is currently, and it's done once a year.

Follow-up question about the prior IRI sample size question for IRI Consumer data – What is the sample size for 2017-18?

For each year, there's going to be 120,000 households responding. You should limit your analysis to 60,000 households usually to make internationally representative inferences limited to the 60,000 that are in the so-called static panel. And that's going to account for probably about each year like 100 million purchases.

Does this mean there is no method of payment info listing EBT or WIC prior to 2017?

That's correct.

National Household Food Acquisition and Purchase Survey (FoodAPS)

Do any IRI or FoodAPS data have geographic location? County level or just state?

Data are geocoded in FoodAPS and TDLinx, and IRI has FIPS.

In the FoodAPS data, is the food away from home (FAFH) data representative of everyone in the household or the individual participant?

For the restricted-use FoodAPS data, the sampling is done at the household level. We do collect information from all the individuals—specifically, we ask that all the food away from home events are reported for all the individuals in the household over the seven-day survey week.

Is there a new wave of FoodAPS data going to be collected since the previous survey?

Right now, planning for FoodAPS-2 is in development. We do have a pilot study that is going to be conducted later this summer that will inform the survey design for FoodAPS-2.

Is FoodAPS able to provide regional data or just national estimates?

The sample is designed for data to be nationally representative. However, geocodes are provided so location information can be used in models.

What nutrient files are included in the FoodAPS data?

Each item reported in FoodAPS (both food at home and away from home) is matched to a USDA food code. These food codes are linked to USDA nutrition databases, which include macronutrients, micronutrients, and food pattern equivalents. The nutrient files are public and available on our website.

The FoodAPS data appears to include both public-use and restricted use-data. How can one determine which variables are in the restricted-use data sets for FoodAPS? Do variables from the restricted use-data have to be included in the proposal?

A list of restricted-use variables is available on the Background page of the FoodAPS website. The appendix of the User's Guide also includes more information on what is included in the public-use files versus the restricted-use files. This grant opportunity includes the restricted-use FoodAPS data.

What's the smaller geographic unit available in FoodAPS and in IRI?

FoodAPS includes geocodes for households and the food acquisition places they visited during the survey week.

For IRI household, you would know their census block. With the stores that they're purchasing, you wouldn't know any geographic information on the individual stores themselves. In the sales data, some data would be at the store level, and you'd have the exact address of the store. Then the other half of the sales data is going to be at the RMA or retail marketing level. That's typically going to be like a region that may span multiple states. But you would know all of the individual stores that make up the automatic. And for all those stores, you would have the address.

For FoodAPS, can a HEI be calculated? What level of dietary data is captured? To the nutrient level? Is estimating quantities necessary to create the HEI?

Yes, HEI scores can be calculated using FoodAPS data. Each food item acquired by households is matched to a USDA food code and our USDA nutrient databases. These USDA nutrient databases include macronutrients, micronutrients, and food pattern equivalents, which allow you to calculate HEI scores. Quantities are included in the data and are necessary for calculating HEI. The FNDDS has about 60 nutrients.

For the Consumer Network, you do not need to have the quantities of the random weight produce items to calculate the HEI scores. But there is the caveat that the vegetable component score and thus the overall score is going to be lower than other measures of HEI. You end up missing about half of the expenditures on produce items. As long as you're careful and making comparisons within the dataset, all the households will be missing the random weight produce items. You'd probably want to do some testing on whatever subgroups or control variables you are using to make sure that one group does not spend a significantly different ratio of random weight produce to packaged produce compared to the other groups. The protein scores seem to line up fairly well with NHANES food at home consumption. But again, you would want to test.

SNAP and WIC Administrative Data and Accessing Data at FSRDCs

Are the administrative data at FDRDCs the only dataset that contains information on SNAP? Is it SNAP eligibility or SNAP use?

The SNAP and WIC state administrative records data include information on participants in the programs. These data do not include information on eligible individuals who never apply for or participate in the programs.

In addition to the administrative records, some Census Bureau surveys ask respondents questions about SNAP and WIC benefit receipt. Examples of surveys that ask questions on these topics include the Current Population Survey Food Security Supplement and Annual Social and Economic

Supplement, the American Community Survey, and the Survey of Income and Program Participation. The FSRDC administrators can tell you more about the various data sources. The slides on the state SNAP and WIC administrative data include links to research that has compared survey response of SNAP to the SNAP administrative records data. See this website for the slides and presentation: Food and Health Research Using the Consumer Food Data System (CFDS) (mathematica.org).

Restricted WIC/SNAP data: (a) Where can we find a list of the variables that are available in the restricted administrative data? (b) Do you foresee more data being added across states or time periods?

The list of variables in the state SNAP and WIC data available for this grant opportunity is located at: Food and Health Research Using the Consumer Food Data System (CFDS) (mathematica.org)

In addition to the states/years of SNAP and WIC administrative data listed on the above website and in the Call for Proposals (in Appendix A), we now have data for the following states/years:

- IL WIC 2020
- KS WIC 2021
- IA SNAP 2021
- MT SNAP 2021
- NV SNAP 2021
- NC SNAP 2021
- SC SNAP 2021
- WY SNAP 2021

When in an FSRDC, are we allowed to access other Census datasets at the center outside of the SNAP and WIC administrative data? If so, would we be allowed to access only publicly available files or can we request access to restricted files (e.g., American Community Survey, Consumer Population Survey)?

Yes, as noted in the Call for Proposals, you can request access of restricted-use Census Bureau decennial census and survey data. The FSRDC administrators can tell you more about the various data sources.

The research proposal should include all restricted-use data the researcher needs to address the research question. The FSRDC Administrator assisting you will be able to help you identify which data can help to address the research question(s).

For the data sets that require Special Sworn Status, do all research team members need to have Sworn Status or can some researchers not go through the process if they are not working directly with the data?

Special Sworn Status (SSS) is required to access data and view any internal tabulations of the data prior to disclosure approval. If you have team members that do not have SSS, they will only be able to see results once they have been approved through the Census Bureau disclosure avoidance review process. Your FSRDC administrator can walk you through this process.

FSRDC data: Per an earlier answer, it sounds like we are only allowed to bring out results as they will appear in a paper / supplementary material. Does this mean multiple individuals on a proposal will be given access? Can team members work together in the FSRDC to work out any analytical challenges that arise? Would they have to be working together at the same center at the same time to do so?

Yes, only results that have been approved for disclosure can be brought outside the FSRDC. Your FSRDC administrator can provide more information about this process.

Researchers on an FSRDC project can work in multiple FSRDC locations and have access to the same project space. For this grant opportunity researchers will need to make arrangements at each of the FSRDCs they will use. They should reach out to the local FSRDC as soon as possible to ask about seat availability and any applicable access fees.

For typical FSRDC proposals, the methods used to clean and analyze the data have to be outlined ahead of time in detail. Will that be the case for this Call for Proposals and/or the later full proposal submission? Will technical assistance be available on the level of detail needed?

The Call for Proposals includes the requirements for the brief proposals and full proposals: viewCfp.do (rwjf.org). You can work with your FSRDC administrators for questions on what needs to be included.

For FSRDC proposals, can I bring in other data to merge with restricted data?

Yes, you can bring in public-use summary statistics into your project space. You can work with the FSRDC administrators to do this and should include this information in the research proposal.

Food Security Supplement (FSS) to the Current Population Survey (CPS)

Is disaggregated race/ethnicity data available in the FSS to CPS?

Yes, there is detailed race and ethnicity level data within the food security supplement.

General Topics

At what level of geographical granularity are food prices available?

Restricted FCBS and EHM data are not available under this grant opportunity because these modules are collected by National Center for Health Statistics (NCHS) and Bureau of Labor Statistics (BLS). As such, access to microdata data are subject to NCHS and BLS approval and process, making their availabilities of the subject to procedures RWJF and ERS could not control.

Consumer Food Data System is a system of disparate sources of data with different contents. If a researcher is looking for the most granular price information, one could look at IRI which has item-level unit prices and quantities. However, IRI data are for food-at-home items only. But FoodAPS provides item-level prices and quantities for both at home and away from home food items for 2012. IRI provides high-frequency and timely data. Both FoodAPS and IRI have rich geographic information as well as does TDLinx, which has geocodes for store locations.

With respect to the FoodAPS and geographic location, the survey is designed to be nationally representative of the continental United States. So, we construct weights to ensure that estimates

constructed with the survey data are nationally representative. We do have different types of data measured and collected at different levels of geographic granularity. And we would discourage using these data for regional analysis at the state level or county level. You can, of course, incorporate the regional/geographic information in models. However, FoodAPS data do not support research focused on a particular region.

What datasets contain information on SNAP, and is it SNAP eligibility or SNAP use? Is there a dataset that reports information on SNAP use/eligibility at the geographical level, and if so, what it the most granular level of information available?

FoodAPS has SNAP participants. But data are only for 2012. Data at FSRDC are at the State level for SNAP and WIC but not for all States. For more information, see the Call for Proposals, the ERS website, and the data webinar on the CFDS website: https://www.mathematica.org/projects/consumer-food-data-system-grant-competition

The Eating and Health Module of the ATUS has SNAP data as well as state identifiers, and it is available for the years 2006-2008 and 2014-2016. The Consumer Network of IRI has SNAP participation information, and WIC, for subset of households. Also, the Food Environment Atlas has SNAP participant and SNAP eligibility information as well at the state level. And the SNAP participants is at the county level.

Does the CFDS include any data on the SNAP-Ed program?

Unfortunately, not.

Is there any restriction for non-US citizens to access the data if the proposal gets accepted, especially with regards to obtaining SSS requirement?

For both ADRF and FSRDC, to be eligible for an award, researchers must be affiliated with U.S. academic organizations that are either institutes of higher education, public entities, or nonprofits that are tax exempt under Section 501(c)(3) of the Internal Revenue Code and are not private foundations or Type III supporting organizations. For FDRCS SSS, more specific eligibility criteria are detailed in Appendix B of the Call for Proposals. Please review it for further information.

Can you clarify which datasets are included in the public-use CFDS?

In Appendix A of the Call for Proposals, there is a separate subsection that lists the public-use data assets (subsection D).

Rather than restrictions on geographic information related to participants, is there information on the geographic source of food items, their ingredients, etc.? Particularly for programs like SNAP, WIC, etc.

There is information in FoodAPS with respect to geography of food sources. IRI, TDLinx provide geocodes for stores. In addition, the Purchase-to-Plate links foods as purchased to food as eaten linked to USDA nutrition information.

Is there any awareness of food component data similar to the FICRCD (food intake converted to retail commodities) or FCID (food commodities intake data), but perhaps more updated?

The Food Intakes Converted to Retail Commodities Databases (FICRCD) provide data for foods consumed in the national dietary intake surveys at the retail commodity level.

The survey foods are converted into 65 retail-level commodities. The commodities are grouped into eight major categories: Dairy Products; Fats and Oils; Fruits; Grains; Meat, Poultry, Fish and Eggs; Nuts; Caloric Sweeteners; and Vegetables, Dry Beans and Legumes.

The Food Intakes Converted to Retail Commodities Databases were jointly developed by USDA's Agricultural Research Service (ARS) and Economic Research Service (ERS) for the following six surveys:

- Continuing Survey of Food Intakes by Individuals 1994-1996 and 1998.
- National Health and Nutrition Examination Survey 1999-2000.
- What We Eat in America (WWEIA), National Health and Nutrition Examination Survey 2001-2002, 2003-2004, 2005-2006, and 2007-2008.

The FICRCD has not been developed for WWEIA data since 2007-08.

However, in 2019, ERS published a COVID-19 Working Paper, Shares of Commodity Consumption at Home, Restaurants, Fast Food Places, Schools, and Other Away-from-Home Places: 2013-16 https://www.ers.usda.gov/publications/pub-details/?pubid=100137

The Working Paper supplemented the 2007-08 FICRCD with imputed approximations for new foods and applied it to the 2013-16 WWEIA data to estimate the at-home and away-from-home shares of major commodity groups.

When we do analysis in a restricted data center, can we link to other publicly available datasets? What are we allowed to bring in? And what results or statistics are we allowed to bring out?
This is a very good question that described nicely in the Call for Proposals. The answer is different for ADRF vs FSRDCs. Any extant data can be taken into ADRF. Any export request from ADRF must be reviewed for disclosure risk to make sure identity and private information of individuals and establishments are not inadvertently or directly released.

For proposals that involve working in FSRDCs, you can propose to bring in publicly available summary statistics. For FSRDC researchers, the only restricted-use data that may be requested under this CFP in addition to the state SNAP and WIC data are surveys and censuses of the U.S. Census Bureau. This limitation to Census Bureau surveys and censuses is not driven by FSRDC or Census Bureau policies. In general, a proposal can be submitted to an FSRDC that requests importing data into the FSRDC environment or accessing a wider array of data than those permitted under this CFP.

The greater restrictiveness of the CFDS Call for Proposals for Food and Health Research is motivated by the goal of having final research results ready for sharing at a final conference to take place during the grant period.

For bringing results out of the FSRDCs, results will need to be approved through the Census Bureau's disclosure review process. Your FSRDC administrator can walk you through this process. It may be useful to discuss this with them early on to understand the criteria your output will need to meet.

Can we download ADRF datasets to work on our own computer or how these datasets are accessed?

No, data are confidential and sensitive as are the resulting research. They cannot be exported. All work will have to be done behind the firewall and inside the data Enclave. Any export request must be reviewed for disclosure risk. Please see the announcement for details.

If there is a grad student working on this, will ADRF access be possible for multiple individuals and non-PIs? And accessing the ADRF is via a browser, it's not required to go to a physical location for access, right?

Access to ADRF is covered for the PIs and co-investigators. Access to ADRF is web-based and does not require site visits. Further information is available on the Coleridge Initiative website about their Administrative Research Data Facility.

Is this Call for Proposals targeting research questions that use more than one dataset? Or will proposals that only use one dataset still be competitive (especially if no data can be brought in to FSRDCs)?

The competitiveness of proposed research does not rest with use of multiple data sources. It's more a function of analyzing critical food programs, policies and issues facing the Nation. Feasibility of proposed research using CFDS data assets in Appendix A is also important. Proposals that can better analyze important issues of interest to contribute to the literature using multiple CFDS data sources will be preferred. The announcement discusses and provides details of the evaluation criteria.

If we get accepted to use one dataset and later find out we need another restricted use dataset can we request for additional access?

Yes, if your question remains the same. One important selection criterion is the importance of the policy and program relevance of the research proposed. However, please note that another evaluation rubric is the feasibility of the research proposed.

To verify, the TPA forms for ADRF data, will have the section of USDA/ERS/IRI blank, correct? Since we do not need to get in contact with ARDF before submission?

This is correct, the Brief Proposal submission should include TPAs that are signed by your institution with blank USDA/ERS/IRI cells. No contact with ADRF is required until the awards are made. It is important to note that the language in the TPAs is vendor required language that cannot be changed.

Is there any likelihood that this Call for Proposals will be available again if we miss it this round?

No decision has been made at this point for a second round.

Do we need to receive a grant to apply for access to the gated datasets? Or is there a way for researchers to access this data outside of this Call for Proposals?

If the data are public-use data and available on our website (which are identified in Appendix A of the announcement), you can just basically download and use them. To be qualified for this Call for Proposals, however, proposed research must include one of our restricted datasets that are also identified in Appendix A. Cost of access to ADRF and FSRDCs is covered outside of the award and by RWJF and ERS. No additional grant is needed to access the data in ADRF or FSRDC -- access is

provided to awarded proposals. No additional grants/funding to access the data behind the gates are needed. Please note that the data cannot be exported outside of the ADRF or FSRDCs.

Finally, note the important requirement that, if your research proposal uses commercial data, your respective institution must be willing to sign the Third-Party Agreements (TPAs) with the vendor. The TPAs are posted on ERS website. Let me stress: if you plan to submit a proposal using commercial data, you must indicate with your submission that your institution will sign the TPA as is and without any revisions to the language in the vendor's TPAs.

Are there any plans to expand the current SNAP Policy Database to cover the years past 2016?

An update to 2018 is in process, but a completion date isn't known.