

FoodAPS Webinar Transcription – 11/17/16

Good afternoon everyone and welcome to our webinar, USDA's New Study on Consumer Food Acquisition Patterns or FoodAPS: Debut of Public Use Data. My name is Nancy McNiff and I will be your host today. Our speaker is John Kirklin. John is an economist in the Food Assistance Branch at the Economic Research Service U.S. Department of Agriculture. His research focuses on program evaluation, research design, consumer decisions regarding program participation and food purchasing for the past few years. John has been working on the implementation of USDA's national household food acquisition and purchase survey or FoodAPS for short, the first nationally representative survey of its kind. I think we're ready to start John, so you can now begin your presentation.

Okay, thank you Nancy, and hello everybody. Good afternoon. For those of you on the west coast still good morning. We'll start today with the overview of the structure of today's presentation. I'm going to talk about the FoodAPS survey and then move on to the advantages of being able to access public use files compared to the restricted use files that are presently available. But there are also some limitations of the public use files compared to the restricted use files and I'll go over those. Finally we will talk about the public use file documentation and where the files are, where the code books are and then that'll wrap up the formal presentation and we'll turn it over to questions and answers.

So to begin the overview of FoodAPS. What is FoodAPS? The formal name is the National Household Food Acquisition and Purchase Survey. And you can see there where the acronym comes from. It's a survey that measured all purchases and other food acquisitions reported by all household members over a course of seven days. It paid equal attention to what we call food at home and food away from home, equal attention also was paid to purchase food and free food. For example, dinner at grandma's house or pulling tomatoes out of the garden. One thing to emphasize FoodAPS is not a consumption survey. It's a food acquisition survey and trying to understand what people acquire and how much they pay for it, if they pay at all. It's a nationally representative sample of approximately 5,000 households. We over sampled households participating in SNAP and other low income households. And the final sample includes over 1,500 SNAP households. Nearly 1,200 non-SNAP households with total household income below 185 percent of poverty. And then 2,000 or more households that were not SNAP and had income above 185 percent. The survey was fielded, fielded from April of 2012 to about mid-January of 2013.

FoodAPS is really unlike any other survey. It used diary reporting not recall for all members to improve accuracy. We provided scanners to the household so they could scan barcodes for the foods that they brought home, the groceries. We collected not only expenditures on food but quantities and prices of food, all sources of food were reported both food at home and away from home. And also we used extensive use of existing data. We have proprietary point of sale data that were merged in. We used staff administrative records, EBT transaction records, EBT being electronic benefit transfer, the card that SNAP households use when they want to use their benefits to purchase groceries and all those files were used to reduce respondent burden and to improve data quality.

Here's an example I mentioned that we, households could use scanners to scan in either the UPC from the food that they purchased or one of the booklets that they had, had barcodes in there for foods that typically don't have UPC codes. And also on the far left there, a place type code that they could scan to help so they wouldn't have to write all this stuff down.

As I mentioned there is merged in existing data, price and availability of food items at food stores in the sampled area whether those particular stores were visited during the week long period or not. We also include travel distances and times from home to store for the stores so they went to and the

stores that they reported were their primary and alternate stores that they typically go to. And then finally we have the density and the types of stores that exist within a 1 mile radius of their block group or for households in rural areas a 10 mile radius.

There has already been a lot of research interest in FoodAPS and using the restricted data I listed here some of the publications and journal articles that have come out of FoodAPS but perhaps the other thing to note is this slide shows the number of research projects that are underway based on the FoodAPS data. And we expect and certainly hope with today's release of the public use files that these numbers of researchers using the data will increase dramatically and we'll see more research based on the FoodAPS data.

As far as the actual data collection, there were several data collection steps, it started with a screener interview. Then there were two in-person surveys at the beginning and at the end of the data collection week with the primary respondent, there were food books that each household member over age 10 was asked to maintain so they could write everything down and that they acquired. There were three phone calls during the week with the primary respondent to collect information about who had gone where. And particularly for food away from home, what foods had been acquired. At the end of the data collection period after the final interview, there was a feedback survey where we asked some additional questions. And finally during the week they were asked to fill out a meals and snacks form indicating who ate which meals and snacks during the day.

Here's an overview of the plan data collection week, it shows data collection starting on a Tuesday but that was not you know it was fairly random people started throughout the week on different days. You can see there the three calls for the primary respondent to call in to report the foods that had been acquired since the last contact with the Survey Center.

We did make a distinction during a survey between what we now call food away from home and food at home. Those terms were never used during the actual survey data collection. The booklet that people used, the respondents used had instruction pages in them and those pages identified the types of food or the locations where they could enter information about food away from home and food at home that had been acquired.

So as I mentioned earlier until now all the FoodAPS data have been in restricted use files. In order to use the data, a researcher needed an approved project agreement with ERS, a signed Memorandum of Understanding with the researchers institution. They needed to have annual CIPSEA training and sign a pledge of confidentiality with regard to the FoodAPS data, the restricted data are maintained at the NORC Data Enclave at the University of Chicago. And there are fees associated with accessing those data. Finally all research output needs to be reviewed and approved for disclosure risk before exporting files from the Enclave are allowed. So there's a lot of steps that researchers need to go through to use these data.

So now the FoodAPS data are to become available on the ERS website as public use files. And the advantages to the researcher, ease of access, ease of analysis and use of the analytic results. Files and documentation are posted on the website and researchers can download and use these files and documentation without restriction.

So with regard to the information that is available I've got several slides here breaking it out by the type of information. So here we have household information that is available in the public use files. And there's a wide range of topics that are included.

Here are the individual topics. That is data that are collected at the individual level that are available on the public use files. And I'm not going through all the components. This presentation will be

posted after a while and people can come back in and look at the details.

We will be posting information about food at home and food, food away from home events. When the transaction occurred, who participated, where it took place, the payment amount, number of items purchased and more.

And then also about the items, the food items that were acquired. So we provide all this information about each item that was reported or scanned by the respondent. And with regard to the nutrients there near the bottom, that information was merged in after the survey based on the information provided about the description of the item, in particular, the UPC codes.

So those are some of the advantages of using the public use files. What are some of the limitations?

To protect the confidentiality of respondents, disclosure risk is minimized by withholding some data elements and changing others and the changes include things like recoding, combining variables, top-coding some of the responses and a limited amount of data swapping. I will note that even in the restricted environment, no personal identifiable information is available. So it's not available in either the restricted or obviously the public use files.

So examples of dropped variables, most of the variables indicating location have been dropped and that's location either of the household or the places where they go to acquire food. That means that the public use file data cannot be merged with other data sets based on a geographic identifier. The variables that are kept include the census region, metro/ non-metropolitan status of the household location and whether the household's location is in a rural or an urban census tract.

Other dropped variables include precursor variables used to derive other information such as height and weight have been dropped to determine body mass index the individual's race, the pregnancy status of the individual and the income component amounts have been dropped, all place names have been removed from the files.

There have been some recoding or combining of variables, the individual's relation to the primary respondent, age, education, BMI. We have a number of variables indicating food allergies. Those have been aggregated to the household level. And then finally with regard to dates, some of the dates have been recoded to simply the month in which they occurred but the relationships have been maintained for instance the length of time between the date of last SNAP issuance and the date of a transaction.

Topcoded variables include the number of vehicles in the household, income, non-food expenditures, the number of guests that were there for meals and commuting times and those have been capped at the 99th percentile.

Data swapping, individual characteristics were swapped across households for a limited number of records due to concerns about disclosure risk. We cannot provide specific information about how the data swapping was done however and I can tell you that no food acquisition information or household level information was swapped. The only data that were swapped were from the individual level records.

Okay now with regard to documentation and getting access to the data, there are 11 public use down, different downloadable data files and each one comes in three different formats, SAS, STATA and comma separated values. There's a household file, an individual file for food at home, there is an event file, an item file and a nutrient file and you have those same three files for food away from home. You have a file that indicates the meals and snacks that each household member

reported, you have an access file that provides for instance counts of food outlets by distance. And then there is a file of replicate sample weights.

With regard to documentation, there are 13 files. There is a user's guide which I strongly recommend that you read first. Then there is one codebook per data file except for the replicate sample weights. And that information was covered in the User Guide. Then there are two extra documentation files, one having to do with food away from home and the process that was used with regard to expenditure data for food away from home. And then there was a lot of work that was done to match the food descriptions to the nutrient characteristics. And that is discussed in a document called the Nutrient Coding Overview.

And I mentioned earlier about replicate weights. It's, it's not easy to use either the restricted or the public use data to obtain the variance or the standard error of a variable on the file, there is because it is a complex sample design. We have provided the design information in these three variables respectively. They cover the clustering, and the STRATA and then the weight associated, sample weight associated with the each household or each individual. And you can use that information in many statistical packages to get an approximation of the variance or you can use to replicate weights to get a different perhaps improved estimate of the variance. And this was all discussed in the User's Guide. The important thing to know is that this information is there for those who want to use it.

The public use files can be accessed at this URL. If you click on it you'll land on the landing page which doesn't show up real well here because of the size constraints. But you'll see at the bottom four different files that you can access. The first one are all the SAS data files and then STATA and then the CVP and the last one there you click on it will take it you the code books.

So here's an example. If you click on the SAS data file hyperlink this will come up and you can select whichever file you want and download it.

Here's the document files. Again you can click on and download however many of these that you wish.

And that's it with regard to the data and what they are and what they're not and how you can access them. So I'll turn it over now to you to any questions that you might have.

Yes if you have any questions you can enter them into the Chat Feature and I will ask John the questions and he will he will answer them. We already have a few questions and I was going to ask, were there, will there be another survey like this in the future?

That is always dependent upon funding. We are... We have... We expect to have a second survey, we are in the process of preparing for it now. If you hear a rumble in the background, my apologies that's a train passing by. But yes, there will be a second FoodAPS survey that we'll be working on in the near future.

And does a survey cover food consumption at all?

99 percent, no, this is a food acquisition survey. What was purchased or obtained and how much was paid for it? There was a meals and snacks form that asked about which meals and snacks each household member consumed during the week but it didn't ask anything about what was consumed, only whether it was consumed. Now with regard to the food away from home you know if I report that I got a cup of coffee at work on Tuesday you can sort of assume that I probably drank the cup of coffee but we didn't gather information about place weight or thing like that so we don't know

consumption. We just know acquisition.

Okay, we have a question about the food away from home data, does it include prepared foods in grocery stores.

Yes, we asked the respondents to record either on the blue pages or the red pages of their food books information about the acquisition. So if they were uncertain about where to put it, you know they put it in one or the other so sometimes we found records in the, in the blue pages which were for groceries that really suggested it was food away from home and vice versa. But it's all there with descriptions about the food that was acquired and so as an analyst you can make your own decisions about how you want to categorize the acquisitions but the important thing is we asked for everything.

Okay, we have another question, on what level of location information for household is still available in the public use data? Can researchers determine state and county?

No, the only information that is available in the public use file is census tract. I'm sorry census region. Big difference there. And whether it's urban, rural or metro/non-metro. And the reason for that is because the risk of disclosure of confidential information skyrockets when knowledge about where a particular survey respondent resides is provided. So for that reason state and county information is not available.

Okay, on along the same lines, are the dropped variables that you're talking about are they only dropped for the public use files?

That's correct. If you have a research question that needs that sort of information, those variables are still available in the restricted data use at the data enclave.

And if you go to the website, you can go to a page that talks about access to the data and it will walk you through the steps you need to follow in order to prepare and submit a proposal. Basically a project agreement that covers what it is that you want to do and which data you need to do that. And so yes you can still go through and access the public use. I'm sorry, the restricted use data at NORC.

Okay, could you give an example of an individual level data swap?

No, no, and the reason I don't mean to be flippant. The contractor who, who did this for us stressed to us that we need to be really vague about all the individual steps that were taken to do that because of the possibility of some reverse engineering. So for to protect the confidentiality we acknowledge you know the data swapping was done. If you look at the User Guide, there is a section in there that discusses data swapping. But beyond the information that's provided in the User Guide we've been cautioned not to go into greater detail.

Okay great, how were the participants chosen or decided?

You know saying it was a nationally representative sample. The first step was the contractor selected a bunch of primary sampling units based on probability for portion to size sampling and within each of the primary sampling units, 8 secondary sampling units were selected. And all this is just to get down to some small geographic areas where we could send cost efficiently interviewers out within each of those secondary sampling units. We had two sample frames, one from the postal office address based sampling frame and then we also had a frame of SNAP addresses of households with SNAP participants that we used. And so we ended up with two exclusive sample frames and those were used to randomly select addresses within each of the issues.

And could you go over again what the sample size was?

The overall sample size was 4,826 households of which 1,581 I believe were determined to be households with SNAP participants within them and a little over 2,000 were households with no current SNAP participants with total household income above 185 percent of poverty and the remaining were non-SNAP households with income below 185 percent of poverty.

Okay, could you provide a couple of use case examples for the data? How would people use the data?

Well one thing you can, can do is go to our website and we have a list of all the research projects that are currently using the data and we have the title there and so from the titles you can get a pretty good idea of how the data are being used. But we have people who are looking at food security and the variations in food security by different household groups. The key thing that we have in this survey that other surveys don't have is we have information on food choice. And so one of the primary things that you might want to do with these data are look at how does the level of house, of food expenditures and the mix of food expenditures vary by different population groups and by mix you know you're talking about the decision as to whether to buy groceries and take them home to prepare at home or how much should be spent on food away from home which includes you know going out to restaurants for dinner. The, which food items are being purchased, how, how much is being spent on alcohol and beer? How much is being spent on healthy green leafy vegetables? Some people have been looking at the food acquired and creating a measure of the healthy eating index, the USDA measure and looking at how that varies by household characteristics. Within SNAP we have researchers who have been looking at how food acquisition patterns vary over the course of the monthly SNAP issuance cycle. So I mean there's just an amazing number of things that you can do with these data.

Okay, we have a question about what other data sources can you link to the FoodAPS data? Is it possible to link it with NHANES data?

No it is not, even in the restricted use of, we do not have... Like Social Security numbers of any of the household members. So we wouldn't be able to link based on that. We don't have the household address so you couldn't link on that. So that's really we're unable to do that.

Okay, in the survey regarding foods that are talked about that don't have UPC codes, do those have nutrient information as well?

For the most part, yes. So we had you know that our first effort was take the UPC code to identify the food item and then from that to go and find the matching nutrient characteristics, we oftentimes did not have the UPC code or the UPC code that we had did not match anything in our database. So for, for those we used if an item was not scanned we asked the respondent, well first to turn to that booklet because we had a bunch of barcodes that we use to identify food items that typically don't have UPC codes items that are sold in bulk. And so we have a number of food items identified that way in which case we could match to nutrient characteristics. If they still couldn't find a code to scan then we asked them to write down in their food booklets you know what was it that you purchased or acquired? What size was it? And things like that information. And then we also asked them to give us a copy or give us the receipt from either food at home or food away from home. And so for food at home we or the contractor scanned in all those receipts and use the information from the receipt to identify the food item when it couldn't be obtained from a UPC code. Long story short there was a lot of effort put into identifying the proper... you know coming up with a proper identification of the food item and to enable it to be matched to a nutrient database.

And that was difficult because a number of the nutrient databases aren't set up for food as acquired but more for food as consumed in terms of that is after preparation. And the folks here spend a lot of time working on trying to get that correct.

Okay we have a lot of similar questions about the restricted data. So I am going to try to ask you a few of those are the location variables still available in the restricted data?

Yes.

Okay and how difficult is it to get access to their restricted data and what's the procedure for that?

Well as I mentioned the procedure is outlined on our website under... I think there's a access to data link off of that first landing page. And it's a process where the researcher fills out a project agreement form and describes the planned research. We have all the restricted data categorized into seven, seven different groups and...did everybody just lose the picture?

No I think we are okay.

Okay well I lost it because I timed out.

Well let me see.

Okay so you submit a project agreement to ERS. We review it and we make sure that the research isn't asking to do something that's not suitable or allowed with the data for example, so I want to do a state level analysis of something. Well the data aren't configured, they aren't weighted to do state level analyses and so that would not be approved but if it's a research question that can be addressed with the FoodAPS data then after the review we will approve it. The researcher needs to have their institution enter into a Memorandum of Understanding with ERS and that has mostly to do with confidentiality of the data and in steps to insure against inadvertent disclosure of confidential information. Once those two steps are done and it's approved, then the researcher and members of his or her team who wish to access the data will need, will work with us and will online there is a CIPSEA training module so you take the training, sign a confidentiality agreement with the ERS and then we would provide you the information to contact NORC. And as I said there are there is a fee for, for usage for accessing the data and you would be responsible for that. They have a confidentiality form and they have a tutorial in terms okay here's how you access the data, use the data and request export of any findings. So that's the process, it probably takes at least a month to go through all those steps where you can actually get access to the data, access is web based now.

It used to be through a separate remote terminal but that, they've improved it. So you can do it from your own computer. Those are the steps.

Okay along the same lines, how long does it usually take to get approval?

Well as I said you know I think a month would probably be about the most optimistic. Just because there are a number of layers of review and, and communication, back and forth between ERS and the researcher kind of you know clarify exactly what is it that you're trying to do. You've asked for access to you know one group of data that your project agreement really it doesn't seem that you need access to, to that particular set of data files and so you know so we're not going to grant access to those. Some of these data are proprietary data that we have use of through contractual agreements. And so the researcher would have to enter into a third party agreement with the proprietor to have access to those data and that can take a little bit of time. So it varies a lot. But I think you can probably... We work very hard to try to turn things around as quickly as we can.

Okay here's a question about dropped variable's you were discussing earlier. You mentioned that place names are taking out. Does this include just restaurant names or was this referring to other geographic locations?

Restaurant names have been, been dropped. The type of restaurant, you know there is a categorization but the, the actual names have been dropped from the public use files.

So we have a question about the format of the data that's being released. Will the survey data being made available in SPSS format ever?

No, no there are no plans for that but... so I suspect you could either read it into SAS or STATA and export it in SPSS if that if they do that. Otherwise I mean you can get it export. You can download it as a series of Excel files. And my understanding is that once you've got it in an Excel file or a comma separated format then you could be that into it and make it into an SPSS file.

Okay so here the question about administrative data, is it dropped in the public use files, for example the verified SNAP status?

Well even in the restricted files, researchers can't directly access any of the administrative data. We use the administrative data to create variables like verified SNAP status and those variables are available in the public use file.

Okay can you share how you calculated distance or travel distances and times from home to store for example average time for the week versus day, public transportation versus car, that kind of stuff?

Okay on the file we provide the Euclidian Straight Line Distance between the home and the place of food acquisition. And then we, well the contractor had a subcontractor who, who gathered this information using Google Maps, an application for that. And so you have Google driving distance and time. Now I know that, that can vary depending upon time of day. I suspect that what's there is what was there at the time that the request was made. And then if, if the straight line distance was less than one mile we also captured walking distance and time. And so it's whatever assumptions that Google makes in terms of walking speed that will be built to do that.

Okay for households that were determined to be non-SNAP, is that non-SNAP within the past 12 months or that non-SNAP between, within a smaller period of time?

We define SNAP participation as current SNAP participation. So within the non-SNAP household groups there are...and we asked for people who did not say that they were currently participating. We asked whether anybody in the household had ever participated in SNAP and if they said yes then whether they had participated within the last 12 months that information is available in the public use files.

Okay someone is asking if tobacco purchases were included?

No.

Is it possible to distinguish if they purchase local foods, meaning purchased within the same state?

You have a measure of distance from the household location to store location. So by inference you can largely make that distinction although if you're located on the border of the state you might not be able to capture whether the person had gone across the border to, to buy food.

I think what they're getting at is did you determine whether the food was labeled "Local Food".

No.

Okay with the restricted data can you link by census tracts?

Yes. What you need to do is the variable that you want to merge in you have to upload the census file to NORC and you have to upload the entire U.S. Census file because we don't allow information about where the survey was conducted outside of the Enclave. So you know you had you upload the entire file and then within the Enclave you would do your merge based on the FIPS Code I guess that it has for the census tract. And you have then your merge data.

Okay the next FoodAPS survey that you're planning to do, would it be the same sample, a longitudinal study, or would it be a new sample of people?

It will be a new sample of people because we, we ERS we don't have the original address information. And also you know a number of years will have passed between 2012 and when we could go into the field for FoodAPS 2. And there will be a lot of household mobility and death. You know there were it would be impossible to track down the same people.

Okay is there information about participation in other programs other than SNAP like school lunch program and that sort of thing?

The, on the public use, well for both the restricted and the public use files in addition to SNAP, we asked about WIC participation. We did not match up against administrative data to try to confirm it but we have the report and there's about 400 or so households that said that they had somebody in the household participating in WIC, we didn't ask explicitly about school meal participation. We did determine whether anybody in the household was in school at any point during the data collection week. And if they obtained a meal at school they were supposed to record that as well as the price so that you can you can infer to some...and also you have what was acquired at that event. So you know you can make some differences about whether or not the meal that was acquired was a standard meal for, for breakfast or for lunch and you can determine based on the census tract information. I don't know if anybody's done this but you can probably determine whether the school district is in an area where everybody is eligible for school meals.

Okay as far as the participants were, was there an incentive given to them, and if so how much?

There was incentive. It was kind of a complex incentive structure because we wanted to first you know induce them to participate in the survey. And then once they were participating we wanted to encourage their continued participation throughout the full seven days. First of all for everybody who was, who answered knock on the door for the screener, they received a five dollar bill, to participate everybody who participated, the primary respondent at the end of the week received a 100 dollar check for everybody else in the household if they participated through the week they receive while everybody else in the household, age 11 and above they received a gift card of at the moment I'm blanking. It was like \$25 or \$20 and \$40 or \$10 and \$20. I'm actually I can't remember that. I think it's in the, I'm almost certain it is in the User Guide, but that was to encourage continued participation through the, through the week.

Okay we have a couple questions about the restricted data. Is there a charge for the restricted data for access?

ERS has no charge for access to the restricted data. NORC you know there's a lot of infrastructure

involved in terms of maintaining the Data Enclave and all of the software packages and licenses for all those. So NORC has a license fee of I think it's around \$4,800 per year per project.

Okay and is it necessary to physically go to the NORC facility to access the restricted use data?

No, no, the data can be accessed on the web so you can access it from your home, from your computer, your work computer. Although there are some security requirements that ERS imposes, your, the computer that you use to access it is supposed to be relatively secure. You know you can't have it where co-workers can look over your shoulder and see what you're doing or look in through the window and see the data files that you may be working with.

I think this is sort of a standard protocol. But other than that now you don't have to go to a special facility. You can do it from work.

Okay as far as the non-SNAP or non-SNAP households were they asked why they don't participate in the SNAP program?

No.

Okay did you capture things like whether people carried their lunch or were data coded by restaurant sack lunch etcetera as to what kind of food that they had?

If somebody carried their lunch from home to work or to school, then they shouldn't have filled out anything because they weren't acquiring any food at work or at school. In other words, if the food was taken from the home pantry in the morning and carried to work and consumed there that's not a food acquisition.

Okay, interesting. And another, one more question about the incentives. Did the participants know they were getting incentives by doing a survey when they started their food diary?

Yes they did. And I'll make, I'll note that for instance OMB usually prefers that incentives be given upfront because I think that has a better impact on encouraging participation. But we didn't want the dollar amount of the incentive being used to encourage food participation during the week. So that's why all the incentives were provided during the final, after the end of the final interview except for the \$5 screener.

OK John I think that's all the questions we have for now. Everybody if you have further questions, feel free to e-mail the FoodAPS@ers.usda.gov email address and they will all through your questions. Thank you all for joining us today and have a great day.

Okay thank you all.

Thank you. Please standby