United States Department of Agriculture

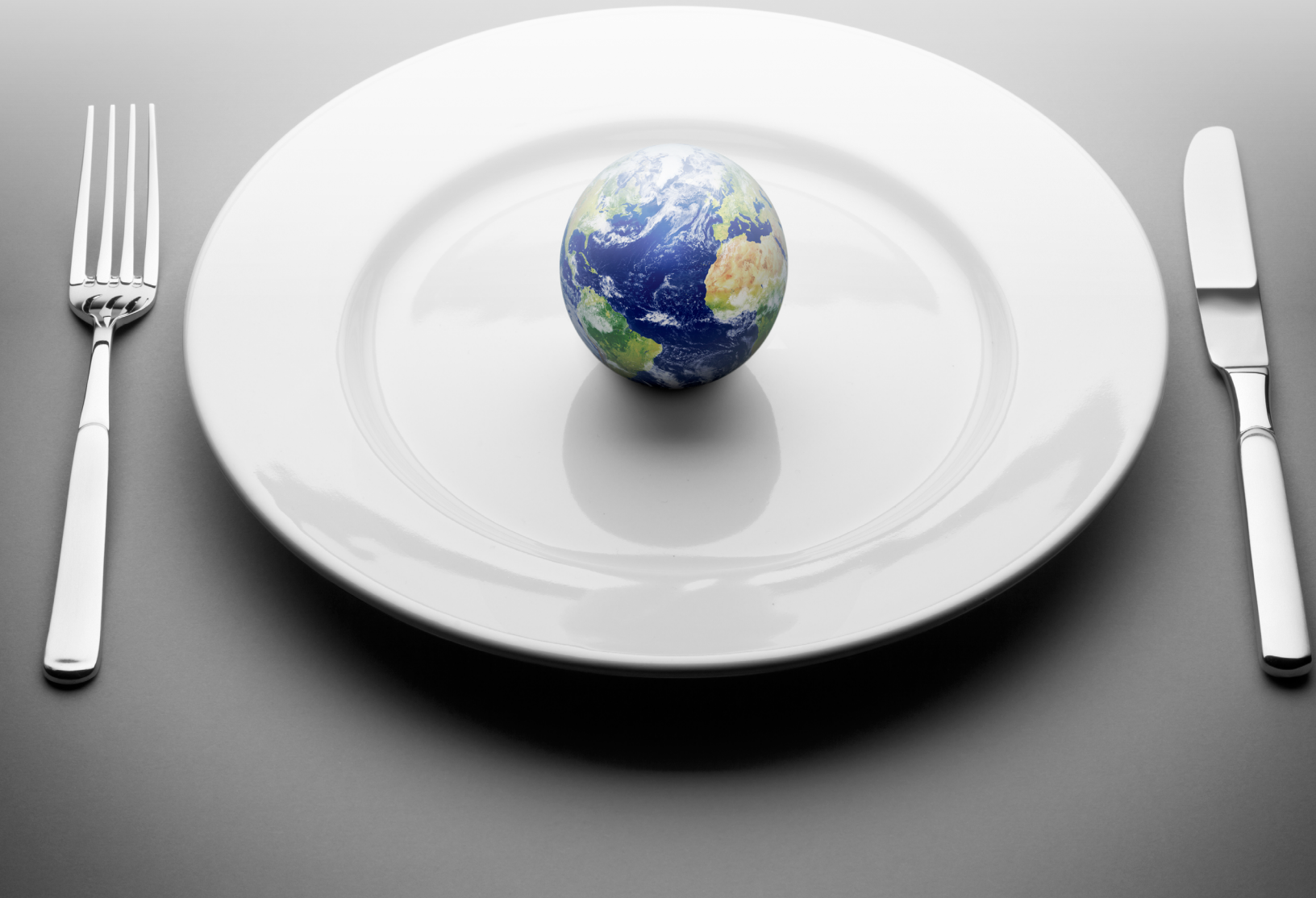
Economic Research Service

Technical Bulletin Number 1953

March 2019

Capturing the Complete Food Environment With Commercial Data: A Comparison of TDLinx, ReCount, and NETS Databases

Clare Cho, Patrick W. McLaughlin, Eliana Zeballos, Jessica Kent, and Chris Dicken



United States Department of Agriculture

Economic Research Service
www.ers.usda.gov

Recommended citation format for this publication:

Clare Cho, Patrick W. McLaughlin, Eliana Zeballos, Jessica Kent, and Chris Dicken. *Capturing the Complete Food Environment With Commercial Data: A Comparison of TDLinx, ReCount, and NETS Databases,* TB-1953, U.S. Department of Agriculture, Economic Research Service, March 2019.

United States Department of Agriculture

Economic Research Service

Technical Bulletin Number 1953

March 2019

Capturing the Complete Food Environment With Commercial Data: A Comparison of TDLinx, ReCount, and NETS Databases

Clare Cho, Patrick W. McLaughlin, Eliana Zeballos, Jessica Kent, and Chris Dicken

Abstract

To study how consumers' diet quality and health may be affected by the food environment, or the number and types of food outlets available in their local communities, USDA's Economic Research Service uses three proprietary datasets: Nielsen TDLinx (food-at-home (FAH) establishments), NPD ReCount (food-away-from-home (FAFH) establishments), and National Establishment Time Series (NETS) (establishments across all industries). This study compares the 2012 data in these three datasets to each other and to the 2012 Economic Census to determine the relative coverage of the food environment across the United States in each dataset. Findings show that NETS reports a higher number of FAH and FAFH establishments operating in the United States than does TDLinx or ReCount, and that the share of TDLinx matches to FAH establishments in NETS is higher than the share of ReCount matches to FAFH establishments in NETS. Findings also highlight the difficulty of including nontraditional FAH retailers, such as drug stores and dollar stores, while excluding establishments that do not sell food. In addition, the findings highlight the importance of using innovative techniques to gather information on FAFH establishments.

Keywords: NETS, TDLinx, ReCount, food at home, food away from home, food environment

Acknowledgments

The authors thank the following individuals for technical peer reviews: Michael Boland, University of Minnesota; Michael Thomsen, University of Arkansas; Alana Rhone, USDA's Economic Research Service (ERS); and TusaRebecca Pannucci, USDA's Center for Nutrition Policy and Promotion. They also thank Tim Park and Constance Newman of USDA, ERS for their comments, and John Weber and Korrin Kim of ERS for editorial and design services. The analysis, findings, and conclusion expressed in the paper should not be attributed to Nielsen, NPD, or Walls & Associates.

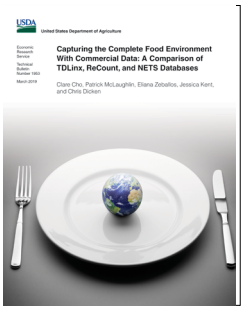
About the authors

Clare Cho, Patrick McLaughlin, and Eliana Zeballos are agricultural economists at USDA, ERS. Chris Dicken is a systems analyst at ERS. Jessica Kent was an intern at ERS at the time this report was written and is currently the financial operations manager at FrontStream Holdings LLC.

Contents

United States Department of Agriculture

A report summary from the Economic Research Service March 2019



Capturing the Complete Food Environment With Commercial Data: A Comparison of TDLinx, ReCount, and NETS Databases

Clare Cho, Patrick W. McLaughlin, Eliana Zeballos, Jessica Kent, and Chris Dicken

What Is the Issue?

Among researchers and policymakers, interest is growing in the relationship between Americans' local food environment, such as grocery stores and restaurants, and their overall diet quality and health. Of particular interest are low-income Americans, who are more likely than other individuals to live farther from grocery stores and to have diet-related health conditions and risks. To facilitate research on the food environment, ERS has purchased three commercial datasets: Nielsen TDLinx (food-at-home (FAH) establishments, i.e., grocery stores), NPD ReCount (food-away-from-home (FAFH) establishments, i.e., restaurants), and the National Establishment Time Series (NETS) (establishments across all industries). These three datasets provide a comprehensive list of food establishments across the United States on at least an annual basis, including detailed information such as geographic locations and sales levels for each establishment.

This is the first nationwide study comparing 2012 TDLinx, ReCount, and NETS to each other and to the 2012 Economic Census (EC) to evaluate the relative coverage of the food environment in each dataset. The EC is considered the official measure of U.S. businesses, but this survey is conducted every 5 years and the publicly available data are aggregated by county and by combined/metropolitan/micropolitan statistical area. Findings can help researchers and policymakers determine which dataset would be most suitable for their data needs.

What Did the Study Find?

Most of the FAH establishments in TDLinx (72.2 percent) match an FAH establishment in NETS. However, only 19.1 percent of FAH establishments in NETS match an FAH establishment in TDLinx; the match rate increases to 31.7 percent when the FAH category is narrowed to grocery stores only. Although the matches between TDLinx and NETS are not confined to a particular geographic area, urban and low-poverty counties (poverty rate 20 percent or lower) had slightly higher shares of matches than rural and high-poverty counties; the difference in matches between urban and rural counties was greater than between high- and low-poverty counties. Part of the difficulty with matching FAH establishments from these two datasets seems to stem from discrepancies in classification, particularly for nontraditional FAH establishments, such as dollar stores, drug stores, and supercenters. In addition, TDLinx does not

classify smaller grocery stores with less than $1 million in annual sales in its grocery trade classification, which may have contributed to the lower share of matches and the higher sales estimates when compared with NETS.

Only 36.3 percent of the FAFH establishments in ReCount match the FAFH establishments in NETS, and only 33.9 percent of FAFH establishments in NETS match the FAFH establishments in ReCount. There is some evidence of regional variation across States: the prevalence of matches are about 2 to 5 percentage points lower than the national average in many States in the West, South, and Northeast. Urban and low-poverty counties in ReCount and NETS had slightly higher match rates in FAFH establishments than rural and high-poverty counties, but high-poverty rural counties had the lowest match rate. Part of the difficulty with matching FAFH establishments from these two datasets is that most restaurants are small, independent operations, which may be less likely than chain restaurants to be part of business registries or have a strong internet presence. In addition, studies show that FAFH establishments experience high turnover rates. Therefore, if an establishment opens and closes within 1 year's time, one data source could report the establishment as open while the other never reports it at all.

Aggregating the establishments in TDLinx, ReCount, and NETS to a county level allowed for a comparison with data in the 2012 EC. In most counties, the EC reports fewer FAH establishments than NETS and TDLinx (93.5 and 63.7 percent of all counties, respectively). In contrast, the EC reports more FAFH establishment than NETS and ReCount (91.3 and 90 percent of all counties, respectively).

How Was the Study Conducted?

All three datasets examined were created using surveys and research conducted by the respective companies, although the specifics on the methodologies used are proprietary. To compare TDLinx, ReCount, and NETS, ERS researchers first geocoded establishments based on street address, city, State, and ZIP Code to obtain a list of all NETS establishments within a one-third-mile radius for each establishment in TDLinx and ReCount. Using this list of potential matches, the researchers matched establishments by name and address using fuzzy matching techniques. Thus, the final dataset contains all matched establishments, as well as unmatched FAH and FAFH establishments in NETS, TDLinx, and ReCount. This dataset was used to examine the matches across datasets and to identify any differences across store classification and location.

Capturing the Complete Food Environment With Commercial Data: A Comparison of TDLinx, ReCount, and NETS Databases

Introduction

There is strong interest in how the local food environment relates to diet quality and health, especially for low-income Americans. Concerns about rising obesity rates and diet-related health conditions have made this an increasingly important policy issue. As a result, Congress requested a USDA, ERS study that examines areas with limited access to affordable and nutritious food and the effect of this limited access on the local population (see Ver Ploeg, 2009). In addition, policymakers introduced programs to incentivize healthy food retailers to enter low-access areas to improve the food environment (e.g., Healthy Food Financing Initiative; Healthy Corner Stores).[1]

Initial studies on the food environment center on geographic access to various food-at-home (FAH) establishments, such as grocery stores, supercenters, and convenience stores.[2] Some studies (e.g., Courtemanche and Carden, 2011; Thomsen et al., 2015) find that the availability of certain FAH establishments can directly affect health outcomes. Others (e.g., Alwitt and Donley, 1997; Caspi et al., 2012; Moore and Diez Roux, 2006; Zenk et al., 2005) find that food environments in geographic areas (e.g., census tracts) with a higher percentage of low-income households and minorities consist primarily of convenience stores and small grocery stores and disproportionately lack access to large supermarkets, suggesting that these differences in food environments may contribute to disparities in health and higher food retail prices. Recently, some studies (e.g., Kyureghian et al., 2012; Pearson et al., 2005; Volpe et al., 2013; Volpe and Okrent, 2012) have also examined the food environment's relation to the healthfulness of food items purchased, particularly fresh fruits and vegetables.

While FAH historically accounted for the majority of U.S. food expenditures and consumption, the prominence of food-away-from-home (FAFH) in the average American diet has grown to rival that of FAH despite a slight decline in FAFH expenditures during the Great Recession (USDA, ERS, 2016; Guthrie et al., 2002; Todd, 2017).[3] The growth of FAFH establishments was primarily due to the increasing number of quick-service restaurants (QSRs—restaurants with counter service rather than table service) in urban areas (McLaughlin and Dicken, 2018). This increase may raise concerns about the diet quality of Americans as some evidence shows that QSRs offer relatively less nutritious foods than FAH and other FAFH establishments (Binkley, 2008; Lin and Guthrie, 2012). Therefore, a number of studies (Alviola IV et al., 2014; Currie et al., 2010; Dunn et al., 2012; Rahkovsky et al., 2018) examine the relationship between proximity to FAFH establishments of various types and dietary or nutrition outcomes. Taken together, findings from these

[1] For more information, see the Healthy Food Financing Initiative page on the U.S. Department of Health and Human Services website and the Healthy Corner Stores page on the Food Trust's website.

[2] We consider establishments where generally food is bought to be prepared and consumed at home or elsewhere as FAH establishments.

[3] We consider establishments that sell prepared food, generally intended to be consumed on the premises of the facility, to be FAFH establishments.

studies suggest that proximity to FAFH establishments may have an influence on dietary and health outcomes in only some geographic and market contexts (e.g., QSRs and nearby schools in the case of Alviola IV et al. (2014)). Thus, research on these and similar issues requires data on the FAFH environment at a granular level.

USDA's Economic Research Service (ERS) has invested in three commercial datasets that have been used to examine the food environment: TDLinx (FAH establishments), NPD ReCount (FAFH establishments), and the National Establishment Time Series (NETS—establishments across all industries). These three datasets provide a comprehensive list of food establishments across the United States on at least an annual basis, including detailed information such as geographic locations and sales levels for each establishment.[4] Some studies (e.g., Gebreab et al., 2017) use Nielsen's TDLinx and NETS databases to examine the effect of the food environment on diet quality. Other studies (e.g., Kaufman et al., 2014; Rahkovsky and Snyder, 2015; Rhone et al., 2017) focus on areas with a large percentage of low-income individuals and low access to supermarkets as identified by the ERS Food Access Research Atlas, which is constructed using TDLinx data matched to an administrative list of stores authorized to redeem USDA Supplemental Nutrition Assistance Program (SNAP) benefits—the USDA Food and Nutrition Service Store Tracking and Redemption System (STARS). This dataset is also listed as a resource that can be used by community leaders applying for a grant under the Healthy Food Financing Initiative. In addition, McLaughlin and Dicken (2018) use NPD ReCount to examine changes in the number of FAFH establishments from 2000 to 2015.

To our knowledge, relatively few studies attempt to validate any of these datasets, most of which focus on a few counties. Rummo et al. (2015) conduct direct field observations in 31 of the 60 census tracts in Durham County, North Carolina, and find that 64 percent of the food stores in the area were included in TDLinx while 55 percent of the food stores were included in Dun & Bradstreet's data. Ma et al. (2013) compare data in the ERS Food Access Research Atlas and the Centers for Disease Control and Prevention (CDC) Non-Healthier Food Retail Tract with data from Dun & Bradstreet and InfoUSA in 169 census tracts across 8 counties in South Carolina. Based on their findings, the authors suggest that these secondary data sources may be acceptable for large-scale surveillance, but they advise local communities to conduct their own fieldwork before implementing efforts to improve food access. Neumark et al. (2005) compare employment levels in California reported in NETS to those in the Quarterly Census of Employment (QCEW), Current Employment Statistics (CES), and the Size of Business; they find that although NETS is generally reliable, the data are initially imputed for new establishments and considerably rounded for employment estimates. Kunkle (2011) compares employment dynamics in NETS to those in the ES-202 unemployment insurance filings (e.g., what underlies the QCEW and CES) and shows that during economic expansions and contractions, NETS may be better at capturing employment fluctuations. Barnatchez et al. (2017) conduct a national study comparing data in NETS to data in three government sources: QCEW, County Business Patterns (CBP), and Nonemployer Statistics. The study examines the number of establishments and employees and by industry and finds that the number of small establishments exhibit the largest differences. Levin et al. (2018) provide some information on TDLinx and NETS but focus on comparing both datasets to the IRI InfoScan Data.

This ERS study is the first nationwide study to compare three commercial datasets often used to depict the food environment: TDLinx, ReCount, and NETS. We compare the FAH establishments

[4] TDLinx provides monthly data, while ReCount provides data biannually; NETS provides annual data.

in TDLinx to those in NETS, the FAFH establishments in ReCount to those in NETS, and the FAH and FAFH establishments in NETS to those in TDLinx and ReCount, respectively. We then conduct a county-level comparison of the FAH and FAFH establishments in each dataset to those in the Census Bureau's Economic Census. Similar to other studies comparing ERS proprietary data to data in other government sources (e.g., Sweitzer et al., 2017; Levin et al., 2018), the results from this study can help researchers determine which dataset would be most suitable for their needs.

Data Description

This report uses FAH and FAFH establishment-level data from three commercial databases: NETS, TDLinx, and ReCount. These databases provide a comprehensive list of food establishments and detailed information about each establishment, including exact geographic location (i.e., latitude and longitude) and sales. Given the increasing share of total U.S. food sales at nontraditional FAH retailers (USDA, ERS, 2017), we include the following under FAH establishments: grocery stores, supermarkets, supercenters, drug stores, and convenience stores. Under FAFH establishments, we include quick-service restaurants and full-service restaurants (establishments with wait service and other amenities).[5]

NETS

Dun & Bradstreet works with Walls & Associates to create NETS, a time-series database that uses Dun & Bradstreet's archival data to provide annual data, with information dating back to 1990 (Walls & Associates, 2013). It reports two establishment names: Company and TradeName. While every observation has a Company, not all have a TradeName. If an establishment has both, TradeName is most likely to be the storefront or banner name while Company typically refers to the legally licensed name of a business (e.g., CVS is the TradeName and CVS Pharmacy INC is the Company name). NETS also includes information on the store's location (including street address and geocodes), geographic information about the headquarters, and other store characteristics, including industry classification, employment, and sales.

NETS contains the street address of each establishment and its headquarters. It also provides the latitude and longitude coordinates and county Federal Information Processing Standards (FIPS) codes for each establishment, as well as the number of establishments under the same headquarters. NETS assigns a unique identification number for both the establishment and its headquarters, allowing users to track establishments under the same parent company.

NETS categorizes establishments using the North American Industrial Classification System (NAICS) and the Standard Industrial Classification (SIC) numeric codes, allowing users to make standardized industry comparisons with other data. These codes are published for each year (e.g., SIC12 for 2012, NAICS08 for 2008) to capture changes in establishment classification over time. While only primary NAICS codes are published, the current primary, secondary, tertiary, fourth, fifth, and sixth SIC codes are published for establishments that engage in multiple business activities; only the primary SIC code is provided for earlier years. The NETS variable Industry is the name of the primary eight-digit SIC code, while IndustryGroup provides the name of the broader four-digit SIC code. For example, an establishment classified as SIC 54110103 has an Industry of "Supermarkets, independent" and SIC 54110101 has an Industry of "Supermarkets, chain," but both are in the IndustryGroup "Grocery Stores." Table 1 lists the four-digit and six-digit FAH and FAFH SIC codes analyzed in this study.

[5] We exclude Dinner Theater (SIC 58129908) and Contract Food Services (SIC 58129906).

Table 1
NETS FAH and FAFH store classifications

SIC	Primary classification	SIC	Secondary classification
5411	Grocery Stores	541101	Supermarkets
		541102	Convenience Stores
		541199	Grocery Stores
5421	Meat and Fish Markets	542101	Fish and Seafood Markets
		542102	Meat Markets
5431	Fruit and Vegetable Markets	543199	Fruit and Vegetable Markets
5441	Candy, Nut, and Confectionery Stores	544199	Candy, Nut, and Confectionery Stores
5451	Dairy Products Stores	545199	Dairy Products Stores
5461	Retail Bakeries	546199	Retail Bakeries
5499	Miscellaneous Food Stores	549901	Health and Dietetic Food Stores
		549902	Beverage Stores
		549999	Miscellaneous Food Stores
53119901	Department Stores, Discount		
53310000	Variety Stores		
53999903	Country General Stores		
53999906	Warehouse Club Stores		
55410000	Gasoline Service Stations		
55419904	Gasoline Service Stations With Convenience Stores		
59120000	Drug Stores and Proprietary Stores		
59129901	Drug Stores		
59129902	Proprietary (Non-Prescription Medicine) Stores		
5812	Eating Places	581201	Ethnic Food Restaurants
		581202	Ice Cream, Soft Drink, and Soda Fountain Stands
		581203	Fast Food Restaurants and Stands
		581204	Lunchrooms and Cafeterias
		581205	Family Restaurants
		581206	Pizza Restaurants
		581207	Seafood Restaurants
		581208	Steak and Barbecue Restaurants
		581299	Eating Places

Note: National Establishment Time Series (NETS) database provides Standard Industrial Classification (SIC) codes at various levels, but for the purpose of this study, we consider the four-digit SIC code as the primary classification and the six-digit SIC code as the secondary classification. We add 9 eight-digit SIC codes to the list to complement the food-at-home (FAH) environment. FAFH = food away from home.

Source: USDA, Economic Research Service.

NETS annually reports sales and employment levels of each establishment. It reports gross annual sales in dollars and employee counts in integers for each year an establishment is open. Both employee count and sales have an associated annual code to indicate the level of reporting accuracy. These estimates are imputed at the establishment level annually and constructed to be used in time trend analyses, which is one of the main advantages of using this dataset.

TDLinx

The Nielsen Company's TDLinx database provides national FAH retail establishment information using independent research and store surveys (Nielsen, 2010). The database is updated on a continuous basis to provide monthly data. In addition to providing individual store names and the names of parent companies, TDLinx identifies store locations by street address, geocode, and FIPS codes. It also provides information on store characteristics, including whether certain items are sold at a particular establishment (gas, liquor, wine, beer, and pharmaceuticals).

TDLinx has a two-tier classification system for each type of FAH retail establishment. Each establishment is assigned a Channel and a Subchannel according to its annual sales and sales volume of specific items (table 2). For example, a retailer could be classified under the Convenience Store Channel or the Cigarette Outlet Channel, depending on whether cigarette sales make up more than 50 percent of the store's total annual sales (Nielsen, 2010). Grocery stores with less than $1 million in sales annually are not assigned a Channel or Subchannel.

TDLinx reports sales under two separate variables: ANNVOL and SWKLYVOL. ANNVOL provides a range of annual all-commodity-volume (ACV) sales,[6] separated into 19 distinct intervals. SWKLYVOL provides an estimate of average weekly ACV sales, calculated using ANNVOL and store attributes, such as location, Channel, and Subchannel. Nielsen Company models the ACV sales on a monthly basis using the best available information. If there are changes in the estimation method, past sales are not updated, making the ACV sales not time-trendable. Thus, both ANNVOL and SWKLYVOL should be considered an estimate of sales based on the best available data, not an actual retail sales report (Nielsen, 2010).

An important store attribute used to calculate ACV sales is the square footage of the selling area, which is also a variable provided only in TDLinx. This information is derived from the actual property site plans, broker visits, direct retailer contact, and newspaper clippings on grand openings or real estate transfers. If individual store information is not available, the selling area is estimated based on corporate-wide annual reports. TDLinx also provides the number of checkout registers in each store.

TDLinx provides an estimate of the number of employees; however, the dataset estimates the number of full-time equivalent employees, where part-time employees are counted as one-half (full-time employees are counted as one). There is no way to identify the total number of employees or to distinguish between the number of full-time and part-time employees.

[6] ACV includes all products offered at the store. In the Grocery Channel, food sales are separated from nonfood sales only for the Supercenter Subchannel, which has an estimate of sales from the grocery store section to make it comparable to the other stores in the Grocery Channel.

Table 2

TDLinx FAH store classifications

Primary classification	Secondary classification	Description
Grocery	Conventional Supermarket	Annual sales of at least $2 million
	Limited Assortment	Limited selection of items in a reduced number of categories
	Natural/Gourmet Foods	Primarily offers natural, organic, or gourmet foods
	Supercenter	Full-line supermarket and full-line discount merchandiser under one roof
	Cash & Carry Warehouse Store	Limited service that concentrates on price appeal
	Military Commissary	Operated within a military installation
	Superette	Annual sales ranging from $1 million to $2 million
Drug	Conventional	Prescription pharmacy items and health and beauty care products
	Rx Only & Small Independent Drug Store	Less than 15 percent of total revenue from items other than prescription drugs; less than $1 million in annual sales
Mass, General Merchandiser, and Dollar Store	Mass	Typical size 40,000 to 160,000 sq. ft.
	General Merchandiser	Typical size 10,000 to 40,000 sq. ft.
	Dollar Store	Most common price is $1; typical size 3,000 to 10,000 sq. ft.
	Military Exchange	Operated within a military installation
Wholesale Club	Membership club stores distributing packaged and bulk foods and general merchandise; average store stocks 4,000 SKUs	
Convenience Store	Gas Station/Kiosk Conventional	Limited selection of grocery items; must carry at least two convenience products[1]
	Military	Operated within a military installation

[1] These products include toilet paper, soap, disposable diapers, pet foods, breakfast cereal, tuna fish, toothpaste, ketchup, and canned goods.

Note: The primary classification in TDLinx is Channel, and the secondary classification is Subchannel. FAH = food at home. SKU = stock keeping unit.

Source: USDA, Economic Research Service.

NPD ReCount

NPD ReCount provides location and business information on FAFH establishments, with information dating back to 2000. In particular, the data include exact geographic location (i.e., address), business name, open and close dates, several levels of menu- and service-style categorization, and employee counts.[7] The data are collected in March and September each year and are scheduled to be available in July and January, respectively. Although the exact methodology is proprietary, NPD Group compiles the data using a variety of publicly available and proprietary data sources, such as chain directories from business' headquarters, industry literature, and verifications through internet searches and phone calls.

[7] The description of the data comes from the proprietary documentation provided by NPD (NPD, 2015). Other variables collected by NPD but not available in the version acquired by ERS include annual sales data, parent company information, and regional population indicators.

Each establishment in ReCount is classified according to a four-level hierarchical categorization scheme. First, all establishments are categorized under Segment as a full-service restaurant (FSR) or quick-service restaurant (QSR). Under each Segment, establishments are assigned a Group, which captures the broad menu type based on criteria such as the sale of alcoholic beverages, breakfast service, and the prices of menu items (table 3). The final levels of classification are Category and Subcategory, which further describe the type of menu offered. Most of these options describe the cuisine offered at the establishment, such as Asian, or the primary item type offered, such as bagel.

Table 3
ReCount FAFH store classifications

Primary classification	Secondary classification	Description
Full-Service Restaurant	Casual Dining Family Dining Fine Dining	Divided based on alcoholic beverage availability, breakfast service, and menu price points
	Unclassified Restaurants	Establishments that do not clearly fit in any of the three categories and tend to be small and independent
Quick-Service Restaurant	Sandwich/Mexican Snack Specialty	Divided according to menu type

Note: The primary classification in ReCount is Segment, and the secondary classification is Group. FAFH = food away from home.

Source: USDA, Economic Research Service.

Two establishments with the same Category can have a different Segment and Group (e.g., Asian and Mexican), although others are unique to each Segment and Group. This level of categorization is unique to this dataset and provides users with more information about the individual establishments. It is important to note that the unit of observation is based on brand rather than physical location. For example, a single establishment may offer the full menu of both Kentucky Fried Chicken and Taco Bell, where all physical aspects of the business (e.g., employees, cooking equipment, and cash registers) are shared. Instead of counting this business as one FAFH establishment, ReCount will provide two distinct records for each restaurant brand or consider it to be two establishments.

Methodology

To compare the information provided by all three datasets and determine their respective strengths and weaknesses, we link records of all FAH and FAFH establishments operating in 2012 reported by NETS to corresponding records in TDLinx and ReCount and vice versa.[8] We focus on 2012 to compare each dataset to the Census Bureau's Economic Census (EC) as well.[9] Unlike Barnatchez et al. (2017), we match each individual establishment rather than the aggregate across the Nation or by two-digit SIC codes. To combine these three datasets, we first geocoded the establishments before attempting to match the establishments across databases.

Although TDLinx, ReCount, and NETS offer geocodes in their databases, it is unclear which software, coordinate system, and geographic level were used; an examination of the databases suggested that different methods were used across these databases. Thus, we geocoded the establishments in all three datasets based on street address, city, State, and ZIP Code using ArcGIS. Using these new geocodes, we listed all the NETS stores within a one-third-mile radius for each store in TDLinx and ReCount. This allows us to include establishments that had a slight discrepancy in the address, multiple establishments with the same address, or multiple establishments with the same name located close to each other.

Using the list of potential matches within a one-third-mile radius, we attempted to match FAH and FAFH establishments by name and address in NETS to establishments in TDLinx and ReCount, respectively. Because each data source may report an establishment name differently or suffer errors in recording, an exact one-to-one match may not always be obtained by an exact comparison of the establishment name and address. To overcome this problem, we used two so-called fuzzy matching techniques that calculate the "distance" between two strings: SAS functions COMPGED and SPEDIS. These programs use distance functions to calculate the number of character transformations (i.e., substitutions, insertions, or deletions) needed to obtain a string of the same characters in the same order as a specified benchmark string. The distance function of each algorithm is strictly increasing in the number of required transformations, although each assigns different weights to each type of transformation.[10] COMPGED weights each transformation equally (i.e., the Levenshtein distance). SPEDIS uses a weighted editing process that, for example, adds a penalty for transformations that occur at the beginning of a string variable rather than the middle or end and normalizes the distance by the length of the benchmark string.[11] To ensure a robust fuzzy matching process, we used both of these functions to match establishments by name and address.

---

[8] We use ReCount's March 2012 release date to ensure that the annual snapshot would align with the NETS release dates; any restaurant with an open (close) date before (after) March 2012 was removed. We use TDLinx's June dataset because ERS acquires the June data annually.

[9] Data are collected for the EC every 5 years. When this study was conducted, the 2017 EC data had not been released, and ERS had not acquired the 2017 NETS data.

[10] The values returned by each distance function have an ordinal interpretation only. For example, consider that "food" versus "good" and "food" versus "gold," respectively, have distances equal to 1 and 2, according to the COMPGED algorithm, This only indicates that "good" is a better match to "food" than is "gold," according to the criteria of the COMPGED algorithm, not that "good" is twice as good a match as "gold." Likewise, the distance function values from COMPGED and SPEDIS are not directly comparable because each algorithm uses different scoring criteria. The range for SPEDIS distance function includes all integers greater than and equal to zero; the range for COMPGED is positive numbers that are a multiple of 10, including zero.

[11] See Dunn (2014) and Cadieux et al. (2014) for details.

TDLinx and ReCount report one name for each establishment while NETS reports two (Company and TradeName). To keep the names and addresses consistent across datasets, we capitalized all of the letters, removed leading and trailing blank spaces, and removed common prefixes and suffixes that might not be consistent across datasets but provide limited identifying information, such as "corporation" or "corp." We ran COMPGED and SPEDIS for names and addresses separately, running SPEDIS twice to account for the function's sensitivity to the ordering of the variables due to the normalization described above. After selecting the lowest score for names and addresses, we summed the two scores and selected the NETS establishment with the lowest score, or the one with the best matching score.[12] In addition, if all of the NETS establishments had a sufficiently high COMPGED score and a sufficiently high SPEDIS, the corresponding TDLinx and ReCount establishments were considered to not have a match in NETS.[13] Thus, establishments with a slight difference in name and address were considered to be matches but not those where the smallest distance was sufficiently large. Although we may have a few incorrect matches, it allows us to include stores that may not exactly match due to an entry error in either dataset. Our final dataset contains all matched establishments, as well as the unmatched FAH and FAFH establishments in NETS, TDLinx, and ReCount.

[12] In general, matched pairs have low values of the distance functions of both algorithms. Therefore, match candidates with a low sum of distance values tended to have low values of the distance function in each individual algorithm.

[13] Little formal guidance exists in calibrating the cutoffs that define a match or a failure to match because of the ordinal property of the distance functions. Therefore, cutoffs must be chosen in an ad hoc way that balances the frequency of false positives and false negatives. That is, if the chosen cutoff is too low, then we may include an unacceptably large number of matches that are objectively incorrect. However, if the cutoff is too high, we potentially exclude many matches that are objectively correct. We calibrated the cutoffs based on manual inspections of the data until we thought we captured the most matches with minimal, but not zero, incorrect matches. The calibrated cutoffs were a score of 340 or more for COMPGED and 38 or more for SPEDIS.

Results

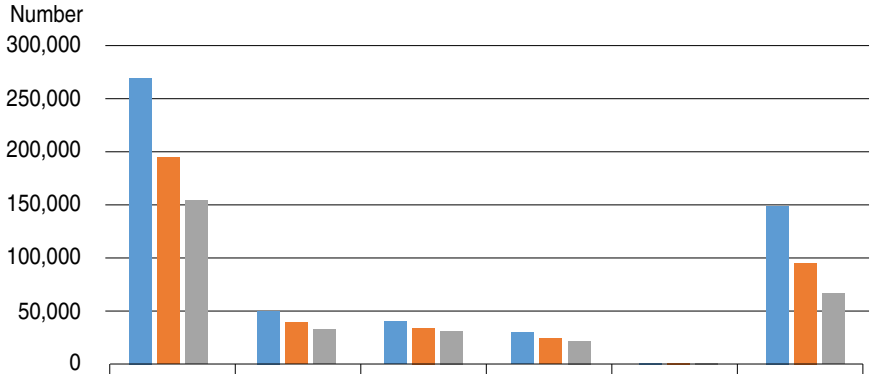
The following presents the results of matching establishment records from TDLinx or ReCount to NETS and vice versa. In general, NETS reports a higher number of FAH and FAFH establishments operating in the United States in 2012 than does TDLinx or ReCount. TDLinx has a higher share of establishments matching to NETS than does ReCount. However, relatively few food establishments in NETS match to TDLinx, while NETS and ReCount have roughly the same match rate regardless of matching direction. Finally, we compare county-level establishment counts from the 2012 EC to all three commercial datasets. The commercial datasets tend to report higher numbers of establishments in FAH and fewer numbers in FAFH than does the EC. In addition, some geographic patterns in matching emerge.

TDLinx establishments in NETS

Among the 269,674 FAH establishments in TDLinx in 2012, 194,693 (72 percent) match an establishment in NETS (fig. 1; table 4). Under TDLinx primary classifications, Convenience Stores have the fewest matches at 64.3 percent, while Wholesale Clubs have the most at 92.8 percent (table 4). The difference in matches could be partially due to the total number of establishments within each classification. While there are almost 150,000 Convenience Stores in TDLinx (over 50 percent of FAH establishments), there are only 1,236 Wholesale Clubs (0.5 percent of FAH establishments). Thus, although the percentage of matches is higher for Wholesale Clubs, the number of matches is much higher for Convenience Stores.

Figure 1

Number of FAH establishments in TDLinx matched to NETS, full dataset and restricted by SIC codes, 2012



	Total	Grocery	Drug	Mass Merchandiser	Wholesale Club	Convenience Store
All TDL	269,674	49,922	39,875	29,877	1,236	148,764
TDL in NETS - full sample	194,693	39,490	34,137	24,298	1,147	95,621
TDL in NETS - restricted sample	154,063	33,220	31,176	21,784	1,056	66,827

Note: Restricted NETS sample includes establishments in the following SIC codes: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores. See Methodology section for details on how SIC codes were chosen.
NETS = National Establishment Time Series Database. SIC = Standard Industrial Classification. FAH = food at home. TDL = TDLinx dataset.
Source: USDA, Economic Research Service using data from TDLinx and NETS.

Table 4

Percentage of FAH establishments in TDLinx matched to NETS, full dataset and restricted by SIC codes, 2012

Channel and Subchannel	% of total	Full matches	Restricted matches*
Grocery	18.5	79.1	66.5
Conventional Supermarket	53.6	81.1	70.6
Limited Assort	6.0	60.7	53.6
Natural/Gourmet Foods	5.3	82.8	69.4
Supercenter	7.6	85.6	80.7
Cash & Carry Warehouse Store	1.0	79.8	13.7
Military Commissary	0.4	68.6	56.2
Superette	26.3	76.6	58.8
Drug	14.8	85.6	78.2
Conventional	73.0	88.5	82.7
Rx Only & Small Independent	27.0	77.8	66.1
Mass Merchandiser	11.1	81.3	72.9
Mass Merchandise	14.1	67.5	59.1
General Merchandise	9.0	78.0	68.3
Dollar Store	75.9	84.9	77.0
Military Exchange	1.1	39.3	4.3
Wholesale Club	0.5	92.8	85.4
Conventional Club	100.0	92.8	85.4
Convenience Store	55.2	64.3	44.9
Conventional Convenience	99.7	64.4	45.1
Military Convenience Store	0.3	16.2	3.7
Grand total	100.0	72.2	57.1

Note: * indicates restricted NETS sample, which includes establishments in the following SIC codes: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores. See Methodology section for details on how SIC codes were chosen. NETS = National Establishment Time Series Database. SIC = Standard Industrial Classification. FAH = food at home.

Source: USDA, Economic Research Service using data from TDLinx and NETS.

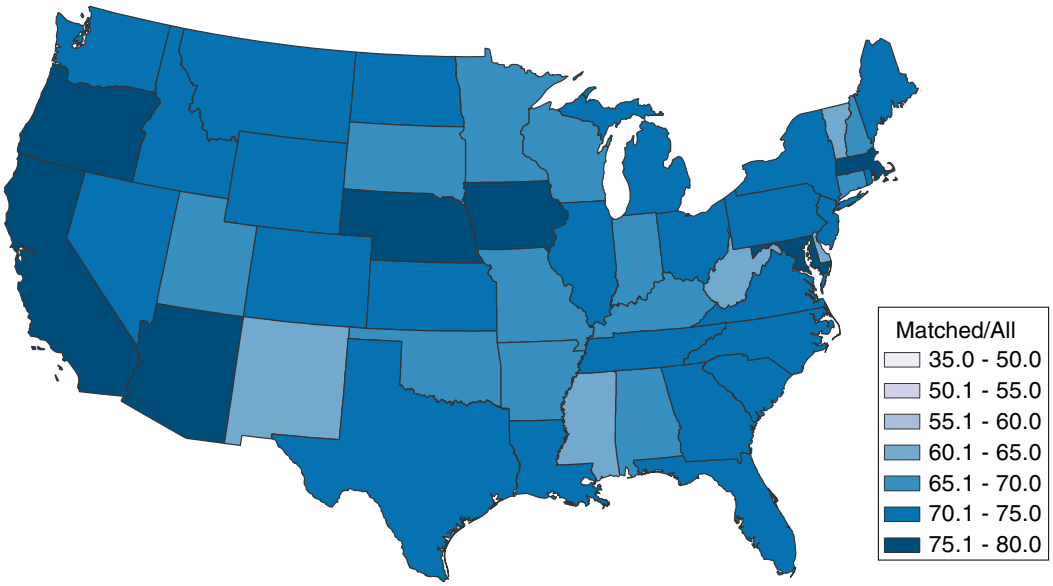
Most of the TDLinx-NETS matches are assigned to similar classifications in both datasets. For example, among Grocery Stores in TDLinx, 78.4 percent match to Grocery Stores in NETS (app. table 1).[14] Establishments in all other TDLinx primary classifications have higher rates of matches to similar NETS classifications, with drug stores having the highest at 97.6 percent. Nevertheless, a few of the matches have an unexpected NETS classification, such as car washes or alarm and safety equipment stores. Thus, we also match a subset of records in NETS restricted to SIC codes for FAH

[14] In addition, among Grocery Stores in TDLinx matched to an establishment in NETS, the second and third most frequent corresponding classifications in NETS were Discount Department Stores and Miscellaneous Food Stores, accounting for 7.6 and 2.5 percent of matches, respectively. Therefore, 88.5 percent of all matches of Grocery Stores from TDLinx have a classification in NETS clearly interpretable as a food retailer.

establishments.[15] Doing so reduces the total number of matches to 154,063 establishments (57.1 percent), with Convenience Stores continuing to have the lowest share of matches (44.9 percent) and Wholesale Clubs having the highest share (85.4 percent) (fig. 1; table 4). However, the share of matches to similar NETS classifications improves using this approach: 82.8 percent of the Grocery Stores in TDLinx match to a Grocery Store in NETS (app. table 1).[16]

Examining the distribution of matches across States illustrates that the prevalence of matches is not confined to a particular geographic area (fig. 2). On average, 70 percent of TDLinx establishments in a given State appear in NETS, with modest deviations for some States. For example, the highest share of matches with the full NETS dataset was in Maryland (80 percent), while the lowest share was in New Mexico and Vermont (63 percent each). Even after restricting the NETS dataset to FAH-related SIC codes, the matches do not appear to be much better or worse in a particular region in relative terms (fig. 3). These findings provide some reassurance that neither dataset appears to be disproportionately underrepresenting the number of establishments at the State level.

Figure 2
Percentage of FAH establishments in NETS matched to TDLinx, full dataset, by State, 2012



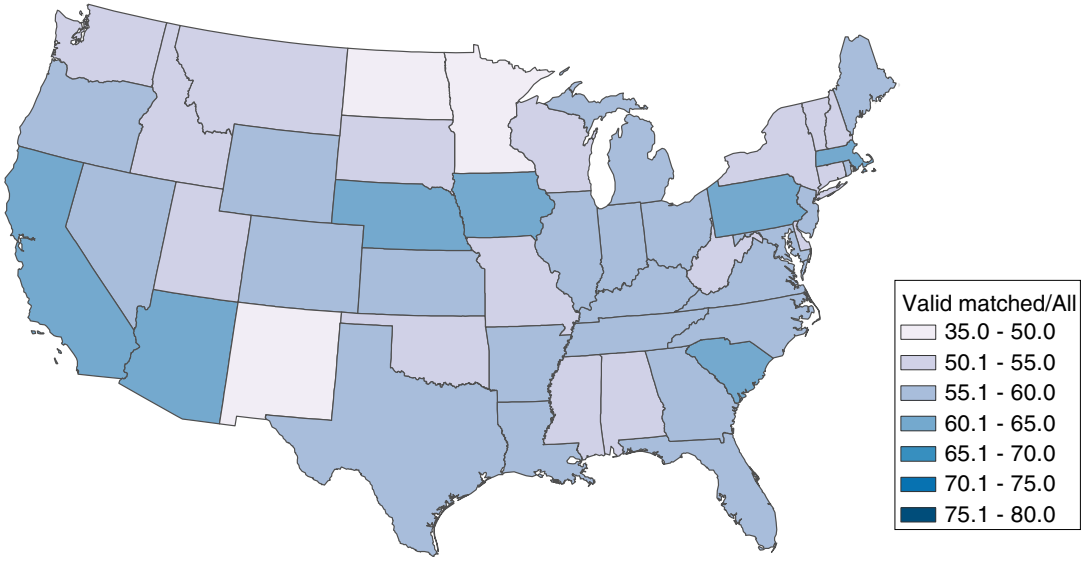
NETS = National Establishment Time Series database. FAH = food at home.
Source: USDA, Economic Research Service using data from TDLinx and NETS.

[15] These establishments are in the following SIC codes: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores.

[16] The percentage of matches to Discount Department Stores and Miscellaneous Food Stores increases to 8.7 percent and 2.6 percent, respectively.

Figure 3

Percentage of FAH establishments in NETS matched to TDLinx, restricted by SIC codes, by State, 2012



Note: The NETS establishments restricted by SIC codes include those in the following: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores. See Methodology section for details on how SIC codes were chosen.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAH = food at home.
Source: USDA, Economic Research Service using data from TDLinx and NETS.

We examined whether the prevalence of matches differed between rural and urban areas, where rural or urban status is defined at the county level using the ERS Rural-Urban Continuum Codes.[17] Although most of the FAH establishments in TDLinx are located in urban counties (77.5 percent), all of the urban counties had better matches than the rural counties by TDLinx classification: 73.5 percent to 67.6 percent for the full dataset, and 58.2 percent to 53.3 percent for the restricted-SIC matches (fig. 4). The only exception was Wholesale Clubs in the restricted-SIC matches, which had 85.4 percent matches in both urban and rural counties. The patterns of rural and urban matches across TDLinx primary classifications were also similar to the overall matches, with Convenience Stores consistently having the lowest matches and Wholesale Clubs having the highest.

We also examined the prevalence of matches by poverty rate. Using the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program, we obtained the poverty rate of each county in 2012 and separated counties into two groups: high poverty (poverty rate greater than 20 percent) and low poverty (poverty rate 20 percent or lower).[18] Comparing the full samples, we found that low-poverty counties had a slightly higher percentage of matches (72.8 percent) than high-poverty counties (70.0 percent); this relationship was maintained for every store type (fig. 5). Even after

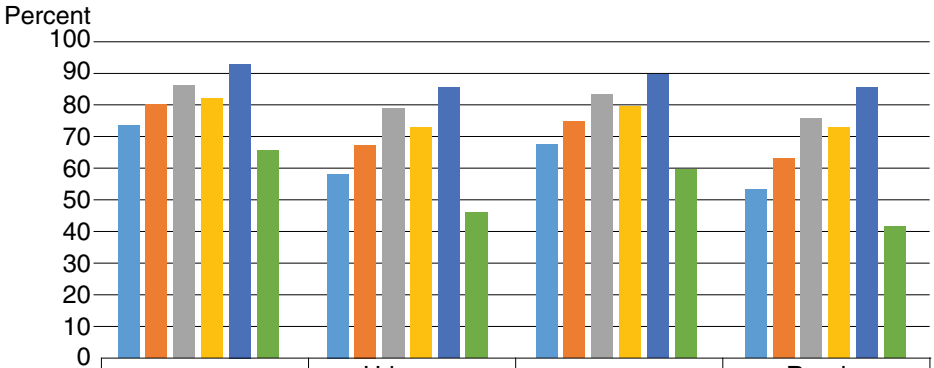
[17] We designated a county coded 1 to 3 as urban and 4 to 9 as rural. See the Rural-Urban Continuum Codes on the ERS website for details.

[18] SAIPE determines the poverty rate by dividing the number of households below the Census Bureau's Federal poverty threshold by the total population. We chose to separate counties by a poverty rate of 20 percent because this is the threshold used by the ERS Food Access Research Atlas and the U.S. Department of the Treasury's New Markets Tax Credit (NMTC) program to determine low-income areas.

restricting the sample by SIC codes, we found the percentage of matches was only slightly higher for low-poverty counties (57.7 percent) than for high-poverty counties (55.2 percent). The difference in the percentage of matches between rural and urban counties was greater than that between low- and high-poverty counties. Among counties with a low poverty rate, the match rate was 73.7 percent in urban counties and 68.8 percent in rural counties; in counties with a high poverty rate, the match rate was 72.9 percent in urban counties and 65.7 percent in rural counties.

Figure 4

Share of FAH establishments in NETS matched to TDLinx, full dataset and restricted by SIC codes, by rural and urban counties, 2012

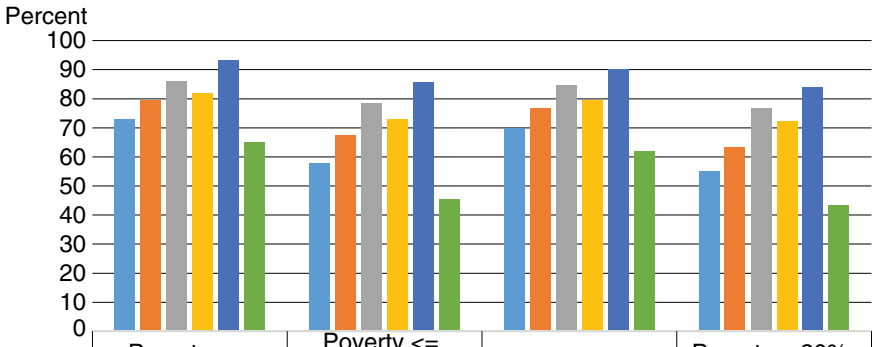


	Urban - full matches	Urban - restricted matches	Rural - full matches	Rural - restricted matches
Total	73.5	58.2	67.6	53.3
Grocery	80.2	67.4	74.7	63.2
Drug	86.1	78.8	83.3	75.6
Mass Merchandiser	82.0	73.0	79.5	72.8
Wholesale Club	92.9	85.4	89.6	85.4
Convenience Store	65.7	45.9	59.6	41.6

Note: Restricted matches include NETS establishments in the following SIC codes: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAH = food at home.
Source: USDA, Economic Research Service using data from TDLinx and NETS.

Figure 5

Share of FAH establishments in NETS matched to TDLinx, full dataset and restricted by SIC codes, by poverty rate, 2012



	Poverty <= 20% - full matches	Poverty <= 20% - restricted matches	Poverty > 20% - full matches	Poverty > 20% - restricted matches
Total	72.8	57.7	70.0	55.2
Grocery	79.7	67.4	76.9	63.5
Drug	85.9	78.6	84.7	76.8
Mass Merchandiser	82.0	73.1	79.5	72.3
Wholesale Club	93.2	85.6	90.1	84.1
Convenience Store	65.0	45.4	62.0	43.3

Note: Restricted matches include NETS establishments in the following SIC codes: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAH = food at home.
Source: USDA, Economic Research Service using data from TDLinx and NETS.

Using the TDLinx-NETS matches restricted by SIC codes, we compare sales information between the two datasets.[19] About half of the matched stores had less than $1 million in sales in NETS, while only 6.7 percent of the matched stores had less than $1 million in sales in TDLinx (table 5). This large discrepancy may have contributed to stores with less than $1 million sales in TDLinx having the highest rate of matches, at 68.9 percent. The majority of matched stores with less than $4 million in sales in TDLinx had less than $1 million in sales in NETS. In fact, there are nearly double the percentage of stores with more than $4 million in sales in TDLinx than in NETS (31.1 percent and 16.6 percent, respectively). Over 50 percent of the matched stores are in a higher sales bracket in TDLinx than in NETS for all but one sales bracket—stores with over $20 million in sales, which also had the second highest percentage of matches in the same sales bracket, at 47.8 percent.[20]

[19] We chose not to conduct the analysis for the matches with the full NETS dataset because the subset restricted by SIC codes provides better matches, as mentioned earlier in this section (app. table 1).

[20] TDLinx does not classify grocery stores with less than $1 million in annual sales under the Grocery Trade Channel. Thus, we conducted the same comparisons mentioned in the paragraph but without the Grocery Trade Channel. We obtained similar results, which are available upon request.

Table 5

Share of TDLinx-NETS matches, restricted by SIC codes, by sales, 2012

TDLinx sales volume	NETS sales volume					Total for TDLinx
	No sales information	<$1M	$1M to $1.5M	$1.5M to $2M	$2M to $4M	
	Percent					*Percent*
<$1M	10.9	68.9	10.3	4.1	4.5	6.7
$1M to $1.5M	14.3	76	5.9	1.8	1.4	22.5
$1.5M to $2M	15.5	69.2	8.8	2.9	2.4	16.8
$2M to $4M	16.1	53.9	12.4	6.2	7.9	22.9
$4M to $6M	13.3	34.2	11.3	10.1	16.7	9.1
$6M to $8M	9.7	15.8	4.8	10.5	20	6.0
$8M to $12M	7.1	8.2	3	8.7	15.2	4.8
$12M to $16M	8.6	7.5	1.7	4.7	10.7	2.2
$16M to $20M	8.4	6.1	0.9	1.6	4.1	1.5
>$20M	7.2	5.4	0.6	0.5	1.3	7.3
Total for NETS	13.2	50.8	7.9	4.7	6.7	

Note: The shaded cells are the percentage of matches that had the same sales volume in both datasets. The NETS establishments restricted by SIC codes include those in the following: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores. See Methodology section for details on how SIC codes were chosen. NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. M = million.

Source: USDA, Economic Research Service using data from TDLinx and NETS.

Table 5 cont.

Share of TDLinx-NETS matches, restricted by SIC codes, by sales, 2012

TDLinx sales volume	NETS sales volume						Total for TDLinx
	$4M to $6M	$6M to $8M	$8M to $12M	$12M to $16M	$16M to $20M	>$20M	
	Percent						*Percent*
<$1M	0.8	0.2	0.2	0.1	0.0	0.1	6.7
$1M to $1.5M	0.3	0.1	0.1	0.0	0.0	0.0	22.5
$1.5M to $2M	0.6	0.2	0.2	0.1	0.0	0.1	16.8
$2M to $4M	1.6	0.5	0.7	0.3	0.2	0.1	22.9
$4M to $6M	5.4	2.0	6.0	0.5	0.2	0.3	9.1
$6M to $8M	12.0	7.5	17.7	1.1	0.5	0.3	6.0
$8M to $12M	13.1	11.0	25.7	4.2	2.3	1.5	4.8
$12M to $16M	8.7	10.2	23.5	12.7	7.6	4.2	2.2
$16M to $20M	8.0	7.7	15.3	18.9	14.8	14.4	1.5
>$20M	3.0	2.3	6.7	12.3	13.0	47.8	7.3
Total for NETS	3.0	1.9	4.3	1.6	1.6	4.0	

Note: The shaded cells are the percentage of matches that had the same sales volume in both datasets. The NETS establishments restricted by SIC codes include those in the following: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores. See Methodology section for details on how SIC codes were chosen. NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. M = million.

Source: USDA, Economic Research Service using data from TDLinx and NETS.

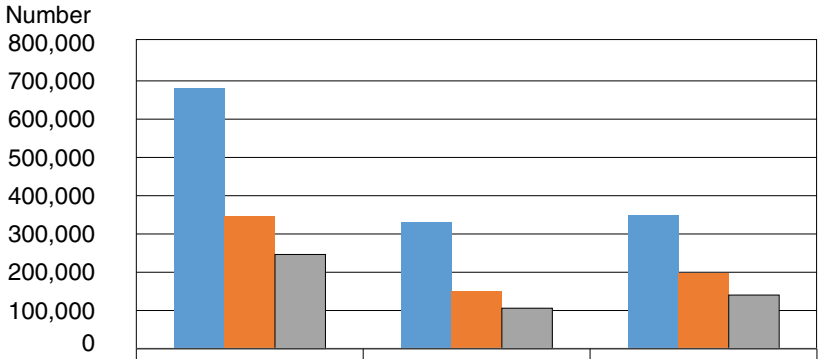
ReCount establishments in NETS

About half (51.2 percent) of the FAFH establishments in ReCount match an establishment in NETS (fig. 6). Although the share of FSR and QSR segments is fairly evenly split in ReCount (49 percent and 51 percent, respectively), the percentage of matches is higher for QSRs (56.7 percent) than for FSRs (45.3 percent). Similar to TDLinx-NETS matches, some of the matches have an unexpected NETS classification, such as Bottled Water Delivery or Youth Organization. Thus, similar to the approach for TDLinx, we also match a subset of records in NETS restricted to only FAFH SIC codes.[21] The share of total matches in ReCount with the restricted-SIC dataset falls to 36.3 percent, with FSRs and QSRs falling to 32.0 percent and 40.3 percent, respectively. Nevertheless, the reliability of the matches improves in terms of consistency of establishments' classifications as defined by both ReCount and NETS. When examining the three most frequent NETS classifications by each ReCount classification, matches to the full NETS dataset range from 80.9 percent to 96.6 percent while matches to the NETS dataset restricted by SIC codes range from 98.4 percent to 99.8 percent (app. table 2).

[21] These establishments are in the SIC code 5812 – Eating Places.

Figure 6

Number of FAFH establishments in ReCount matched to NETS, full dataset and restricted by SIC codes, by primary ReCount classification, 2012



	Total	Full-Service Restaurants	Quick-Service Restaurants
All ReCount	679,252	330,332	348,920
ReCount in NETS - full sample	347,547	149,730	197,817
ReCount in NETS - restricted sample	246,403	105,834	140,569

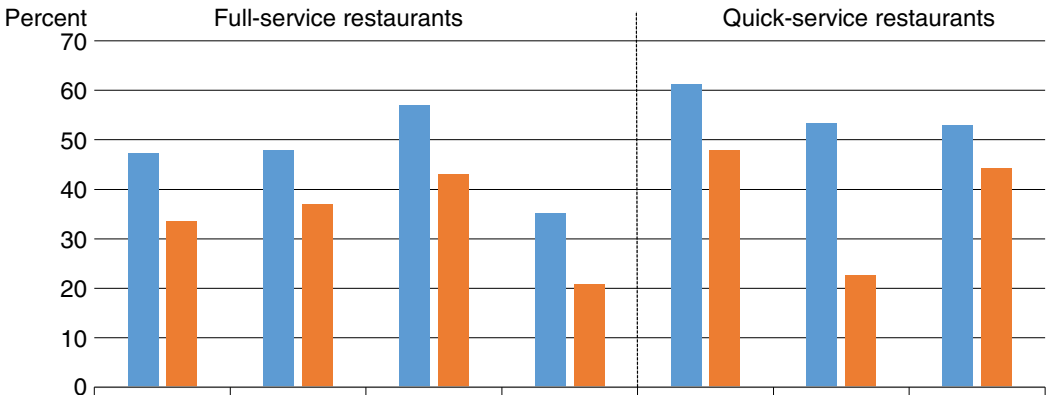
Note: Restricted NETS sample includes establishments in the following SIC code: 5812 – Eating Places.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home.
Source: USDA, Economic Research Service using data from ReCount and NETS.

The ReCount-NETS match rate varies by ReCount's secondary classifications. Among FSRs, Unclassified establishments have the lowest shares of matches to the full and restricted NETS datasets (35.0 percent and 20.8 percent, respectively) while Fine Dining establishments have the highest shares (57.0 percent and 43.1 percent) (fig. 7). Among QSRs, Specialty establishments have the lowest share of matches to the full NETS dataset (53.0 percent). However, compared to the restricted-SIC NETS dataset, Snacks establishments have the lowest share (22.6 percent) and Sandwich/Mexican stores have the highest share of matches to both NETS datasets (61.3 percent and 48 percent, respectively).

The distribution of ReCount matches to the full NETS dataset across States illustrates the regional variation in how well ReCount captures the FAFH environment (fig. 8). In many States, including those in the West, South, and Northeast, the prevalence of matches was lower than the national average of 52.1 percent, mostly ranging from 47 to 50 percent. Nevada, West Virginia, and Vermont have the lowest match rates at 44 percent. ReCount establishments in the Midwest, however, have relatively high match rates, with establishments in Nebraska, Minnesota, Michigan, and Ohio having the highest rates at 57 percent. Figure 9 presents the State-level match rates for the SIC-restricted NETS data. Although the distribution of ReCount matches to the restricted NETS dataset is more evenly distributed across States, most of the Northeastern States had a lower share of matches. Indiana had the highest share of matches (43 percent), while New York had the lowest share (29.4 percent).

Figure 7

Share of FAFH establishments in ReCount matched to NETS, full dataset and restricted by SIC codes, by secondary ReCount classification, 2012

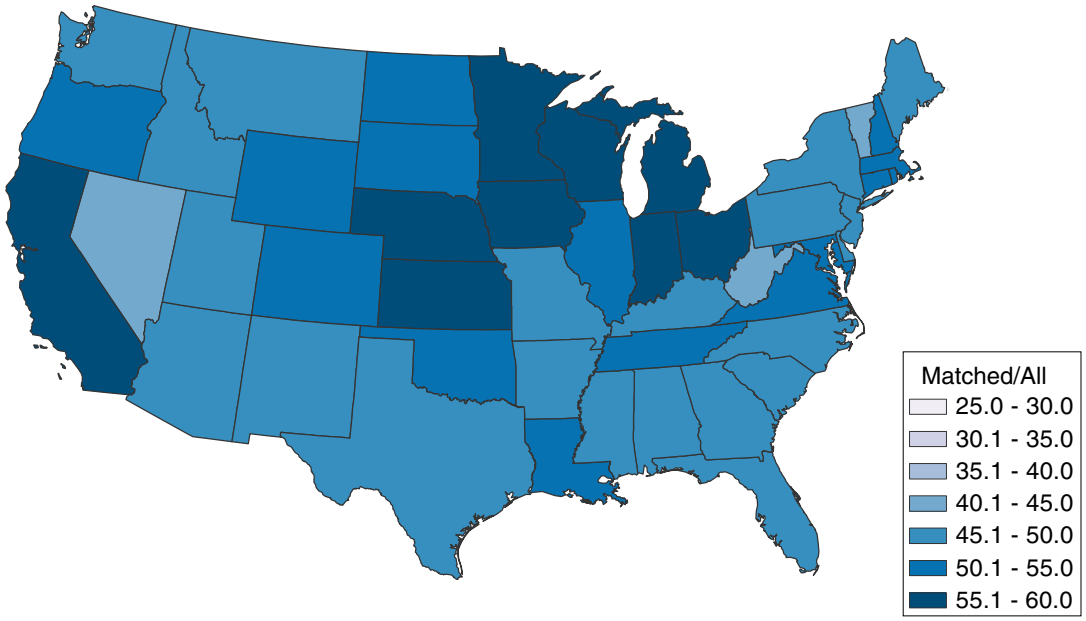


	Casual dining	Family dining	Fine dining	Unclassified	Sandwich/ Mexican	Snack	Specialty
Full matches	47.2	47.8	57.0	35.0	61.3	53.4	53.0
Restricted matches	33.5	37.0	43.1	20.8	48.0	22.6	44.3

Note: Restricted matches include NETS establishments in the following SIC code: 5812 – Eating Places.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home.
Source: USDA, Economic Research Service using data from ReCount and NETS.

Figure 8

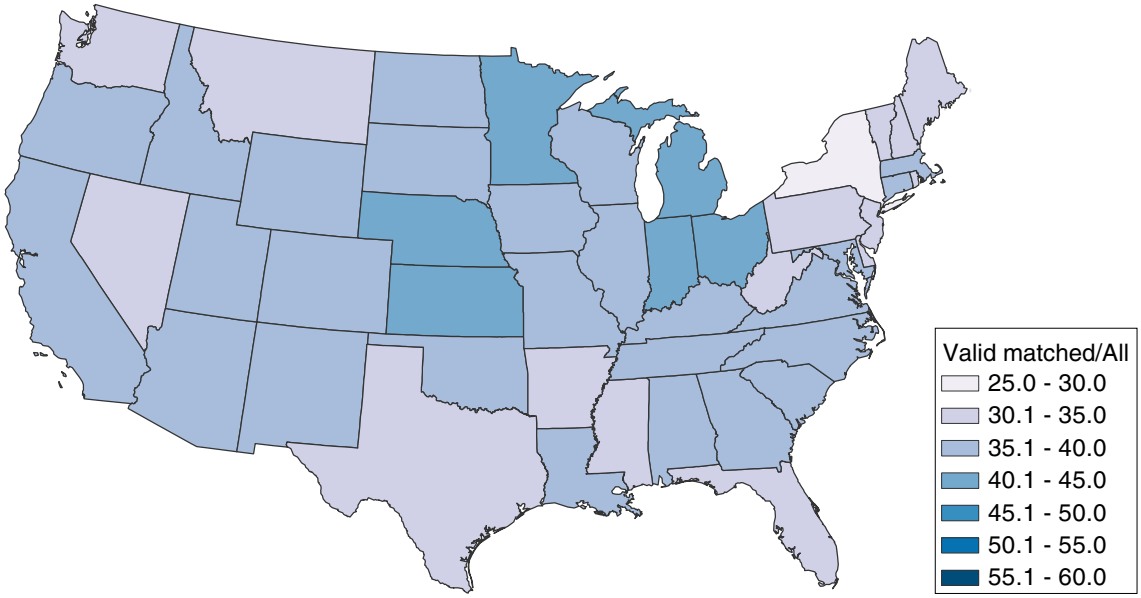
Percentage of FAFH establishments in NETS matched to ReCount, full dataset, by State, 2012



NETS = National Establishment Time Series database. FAFH = food away from home.
Source: USDA, Economic Research Service using data from ReCount and NETS.

Figure 9

Percentage of FAFH establishments in NETS matched to ReCount, restricted by SIC codes, by State, 2012



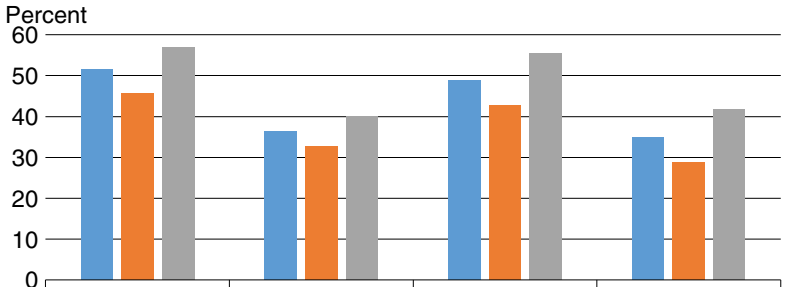
Note: The NETS establishments restricted by SIC codes include the following: 5812 – Eating Places. See Methodology section for details on how SIC codes were chosen.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home.
Source: USDA, Economic Research Service using data from ReCount and NETS.

Most of the FAFH establishments in ReCount are located in urban counties (86 percent). Urban counties had a higher percentage of matches for the full and restricted dataset (51.5 percent and 36.5 percent, respectively) than rural counties (48.8 percent and 34.9 percent) (fig. 10). This is reflected in the matches for FSRs and QSRs in the full dataset; but for QSRs in the restricted-SIC dataset, rural counties have a higher share of matches (41.7 percent) than urban counties (40.1 percent).

There was also a higher prevalence of matches in low-poverty counties (51.4 percent) than in high-poverty counties (49.8 percent) (fig. 11). There was a higher percentage of matches for full-service and quick-service restaurants as well, although the difference for the latter was small—only 0.5 percent. Although the percentage of matches fell when comparing the restricted sample, the overall relationship persisted for all FAFH establishments and for full-service restaurants. However, for quick-service restaurants, the percentage of matches was higher in high-poverty counties (41.0 percent) than in low-poverty counties (40.1 percent). High-poverty rural counties had the lowest percentage of matches at 46.9 percent. High-poverty urban counties and low-poverty rural counties had match rates of 50.9 percent and 49.7 percent, respectively; low-poverty urban counties had 51.6 percent of matches.

Figure 10

Share of FAFH establishments in NETS matched to ReCount, full dataset and restricted by SIC codes, by rural and urban counties, 2012

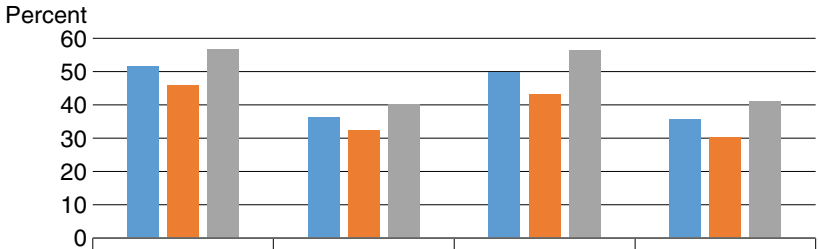


	Urban - full matches	Urban - restricted matches	Rural - full matches	Rural - restricted matches
Total	51.5	36.5	48.8	34.9
Full-Service Restaurants	45.8	32.6	42.7	28.6
Quick-Service Restuarants	56.9	40.1	55.5	41.7

Note: Restricted matches include NETS establishments in the following SIC code: 5812 – Eating Places. See Methodology section for details on how SIC codes were chosen.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home.
Source: USDA, Economic Research Service using data from ReCount and NETS.

Figure 11

Share of FAFH establishments in NETS matched to ReCount, full dataset and restricted by SIC codes, by poverty rate, 2012



	Poverty <= 20% - full matches	Poverty <= 20% - restricted matches	Poverty > 20% - full matches	Poverty > 20% - restricted matches
Total	51.4	36.4	49.8	35.7
Full-Service Restaurants	45.8	32.4	43.1	30.1
Quick-Service Restaurants	56.8	40.1	56.3	41.0

Note: Restricted matches include NETS establishments in the following SIC code: 5812 – Eating Places. See Methodology section for details on how SIC codes were chosen.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home.
Source: USDA, Economic Research Service using data from ReCount and NETS.

NETS establishments in TDLinx and ReCount

NETS had 482,846 Food Stores (SIC 54) in 2012, nearly double the number of establishments in TDLinx. Only 19.1 percent of Food Stores match to TDLinx (table 6). Among the primary classifications, Grocery Stores (SIC 5411) had the highest percentage of matches at 31.7 percent; the matches for other classifications were all less than 4.0 percent. In addition, 7.1 percent of the Food Stores matched to an establishment in ReCount, with 0.3 percent matching to an establishment found in both TDLinx and ReCount. These establishments could be stores that offer both FAH and FAFH, such as a convenience store that offers fast food, or these establishments may have been mistakenly included in one of the datasets.

Table 6

Percentage of Food Stores (SIC 54) in NETS matched to TDLinx and ReCount, 2012

NETS classification (4-digit SIC code)	% of total	Matched to TDLinx	Matched to ReCount	In both	Total matched
Grocery Stores (5411)	58.6	31.7	2.6	0.4	33.8
Meat and Fish Markets (5421)	3.8	3.2	3.4	0.1	6.5
Fruit and Vegetable Markets (5431)	2.8	2.7	0.6	0	3.3
Candy, Nut, and Confectionery Stores (5441)	3.5	0.5	2.1	0	2.6
Dairy Products Stores (5451)	1.5	1.6	27.6	0.2	29.0
Retail Bakeries (5461)	15.1	0.3	19	0	19.3
Miscellaneous Food Stores (5499)	14.6	1.7	14.3	0.1	15.9
Total	100	19.1	7.1	0.3	25.9

NETS = National Establishment Time Series database. SIC = Standard Industrial Classification.

Source: USDA, Economic Research Service using data from TDLinx, ReCount, and NETS.

The other restricted establishments that may provide FAH in NETS match a higher percentage of establishments in TDLinx—35.9 percent (table 7). Variety Stores have the highest percentage of matches (52 percent), while Gasoline Service Stations have the lowest (13.8 percent), likely due to the exclusion of Gas Station/Kiosk in the TDLinx dataset acquired by ERS. Compared to the percentage of matches with Food Stores (SIC 54), fewer establishments match to ReCount (0.4 percent) and match to both TDLinx and ReCount (0.2 percent).

There are 725,684 establishments in NETS under the classification Eating Places (SIC 5812) in 2012, about 46,000 (7 percent) more than the number of establishments in ReCount. A total of 33.9 percent of these establishments matched to an establishment in ReCount (table 8). Fast Food Restaurants and Stands has the highest percentage of matches (48.3 percent), followed by Pizza Restaurants (43.9 percent), while Eating Places, Not Elsewhere Classified (NEC) has the lowest percentage of matches (18.2 percent). In addition, across all the establishments under Eating Places (SIC 5812) in NETS, 0.4 percent match to TDLinx and 0.1 percent match to both TDLinx and ReCount. As mentioned above, this small percentage could be capturing stores that offer both FAH and FAFH or errors in one or all of the three datasets.

Table 7

Percentage of other FAH establishments in NETS matched to TDLinx and ReCount, 2012

NETS classification (8-digit SIC code)	% of total	Matched to TDLinx	Matched to Re Count	In both	Total matched
Department Stores, Discount (53119901)	7.4	27.9	0.5	0.1	28.3
Variety Stores (53310000)	21.7	52	0.3	0.1	52.2
Country General Stores (53999903)	2.2	14.7	0.8	0.1	15.4
Warehouse Club Stores (53999906)	1.3	45.7	0.2	0.1	45.8
Gasoline Service Stations (55410000)	23.7	13.8	0.8	0.2	14.3
Gasoline Service Stations With Convenience Stores (55419904)	0.2	22.4	2	0.7	23.7
Drug Stores and Proprietary Stores (59120000)	20.6	41.7	0.3	0.2	41.8
Drug Stores (59129901)	21.9	43.7	0.2	0.1	43.7
Proprietary (Non-Prescription Medicine) Stores (59129902)	1	16.1	0.4	-	16.5
Total	100	35.9	0.4	0.2	36.2

NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAH = food at home.

Source: USDA, Economic Research Service using data from TDLinx, ReCount, and NETS.

Table 8

Percentage of Eating Places (SIC 5812) in NETS matched to TDLinx and ReCount, 2012

NETS classification (6-digit SIC code)	% of total	Matched to ReCount	Matched to TDLinx	In both	Total matched
Eating Places (581200, 581201)	36.9	27.1	0.5	0.1	27.5
Ice Cream, Soft Drink and Soda Fountain Stands (581202)	3.4	33.4	0.6	0.1	33.9
Fast Food Restaurants and Stands (581203)	24.9	48.3	0.6	0.2	48.8
Lunchrooms and Cafeterias (581204)	1	25.6	0.8	0.1	26.3
Family Restaurants (581205)	7.5	33.8	0.3	0.1	34
Pizza Restaurants (581206)	10.1	43.9	0.5	0.1	44.2
Seafood Restaurants (581207)	1.5	36.2	0.1	0	36.3
Steak and Barbecue Restaurants (581208)	3.2	29.2	0.1	0	29.3
Eating Places, NEC (581209)	11.7	18.2	0.3	0.1	18.4
Total	100	33.9	0.4	0.1	34.3

Note: NEC = not elsewhere classified. NETS = National Establishment Time Series database. SIC = Standard Industrial Classification.

Source: USDA, Economic Research Service using data from TDLinx, ReCount, and NETS.

County-level comparison with Economic Census

We collate the FAH and FAFH store-level information provided by NETS, TDLinx, and ReCount by county to compare the datasets to the Census Bureau's Economic Census (EC).[22] The EC is considered to be the official measure of businesses in the United States, although the survey is not sent to most very small firms to reduce the burden on businesses.[23] We calculate the ratio of establishments in NETS, TDLinx, and ReCount to those in the EC. Thus, if the ratio is smaller (larger) than 1, it indicates that the EC reports a greater (fewer) number of stores; the ratio equals 1 if the reported number of stores is the same.

The EC reports fewer FAH establishments than NETS in 93.5 percent of the counties across the United States; about 1.2 percent of the counties have a ratio of NETS-to-EC establishments greater than 3 (fig. 12).[24] The EC and NETS report the same number of stores in only 91 counties, and in 113 counties, the EC reports more FAH establishments than NETS. These differences could result from very small firms not receiving the EC (to reduce the burden on these businesses) or from the NETS sample including establishments that have an FAH SIC code in any of the six levels of SIC codes while the EC only provides the primary NAICS codes. Although the EC also reports fewer FAH establishments than TDLinx for most counties (63.7 percent), only 0.4 percent of counties have a ratio higher than 3, suggesting the difference between the EC and TDLinx is smaller than the difference between the EC and NETS (fig. 13). In addition, there are a higher number of counties where the EC and TDLinx report the same number of establishments and where the EC reports more FAH establishments than TDLinx—493 and 644 counties, respectively (15.7 percent and 20.5 percent of the counties, respectively).

NETS and ReCount report a higher number of FAFH establishments than the EC for only 4.9 percent and 6.3 percent of the counties, respectively (figs. 14 and 15).[25] NETS reports fewer FAFH establishments than the EC in 91.3 percent of the counties. The remaining 117 counties (3.75 percent of all counties) where NETS and the EC report the same number of establishments are spread relatively evenly across the United States. Similarly, ReCount reports fewer establishments than the EC in 90 percent of the counties. There are a similar number of counties—115, or 5.7 percent of all counties) where ReCount and the EC report the same number of FAFH establishments

[22] Although store-level information is collected through the EC, only the aggregated totals are publicly available. We chose to use the EC rather than the County Business Patterns (CBP) because the EC is considered the official count of U.S. businesses. In addition, the CBP relies on the EC to conduct its estimations.

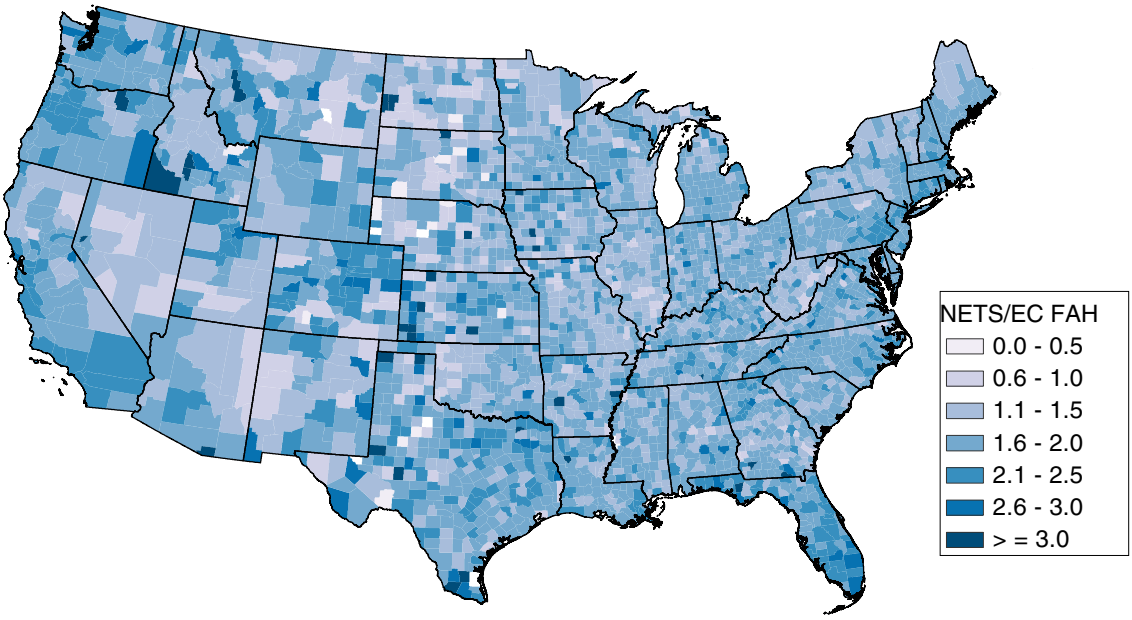
[23] See Economic Census Frequently Asked Questions for additional details. The following NAICS codes are included for FAH: 311811, 445110, 445120, 4452, 446110, 446191, 447110, 452112, 452910; and for FAFH: 722320, 72251, 722511, 722513, 722514, 722515.

[24] The restricted NETS sample in this case includes establishments in the following SIC codes: 54 – Food Stores (except 54210202 – Freezer Provisioners, Meat); 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; and 5912 – Drug Stores and Proprietary Stores; the remaining two SIC codes (53999903 – Country General Stores and 55419904 – Gasoline Service Stations With Convenience Stores) are also excluded. The three SIC codes are excluded from this analysis because we could not find a unique NAICS match.

[25] The restricted NETS sample in this case includes establishments in the following SIC code: 5812 – Eating Places (except 58129908 – Dinner Theater; and 58129906 – Contract Food Services). The two SIC codes are excluded from this analysis because we could not find a unique NAICS match.

Figure 12

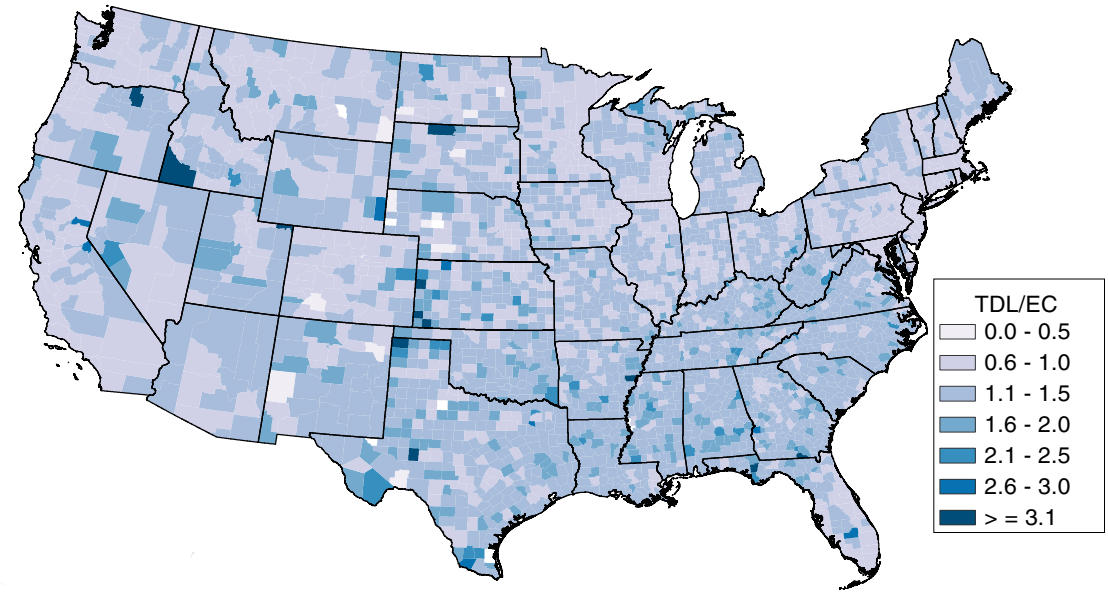
Ratio of FAH establishments in NETS (restricted by SIC codes) to Economic Census, by county, 2012



Note: FAH establishments in the Economic Census are in the following NAICS codes: 311811, 445110, 445120, 4452, 446110, 446191, 447110, 452112, 452910. FAH establishments in the restricted NETS sample are in the following SIC codes: 54 – Food Stores (except 54210202 – Freezer Provisioners, Meat); 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; and 5912 – Drug Stores and Proprietary Stores; the remaining two SIC codes (53999903 – Country General Stores and 55419904 – Gasoline Service Stations With Convenience Stores) are also excluded. The three SIC codes are excluded from this analysis because we could not find a unique NAICS match. Nonshaded counties have no establishments, and a grey county has a ratio of 1.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAH = food at home. NAICS = North American Industrial Classification System. EC= Economic Census.
Source: USDA, Economic Research Service using data from NETS and the Economic Census.

Figure 13

Difference between the number of FAH establishments in TDLinx (TDL) and the Economic Census (EC), by county, 2012



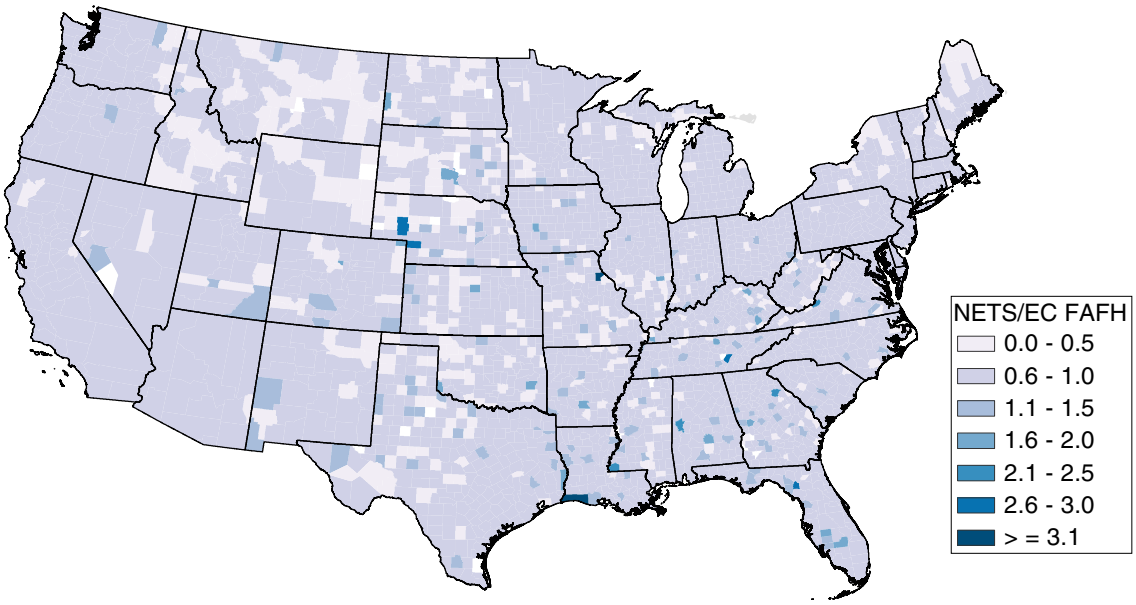
Note: FAH establishments in the Economic Census are in the following NAICS codes: 311811, 445110, 445120, 4452, 446110, 446191, 447110, 452112, 452910. FAH establishments in the restricted NETS sample are in the following SIC codes: 54 – Food Stores (except 54210202 – Freezer Provisioners, Meat); 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; and 5912 – Drug Sores and Proprietary Stores; the remaining two SIC codes (53999903 – Country General Stores and 55419904 – Gasoline Service Stations With Convenience Stores) are also excluded. The three SIC codes are excluded from this analysis because we could not find a unique NAICS match. Nonshaded counties have no establishments, and a grey county has a ratio of 1.

NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAH = food at home. NAICS = North American Industrial Classification System.

Source: USDA, Economic Research Service using data from TDLinx and the Economic Census.

Figure 14

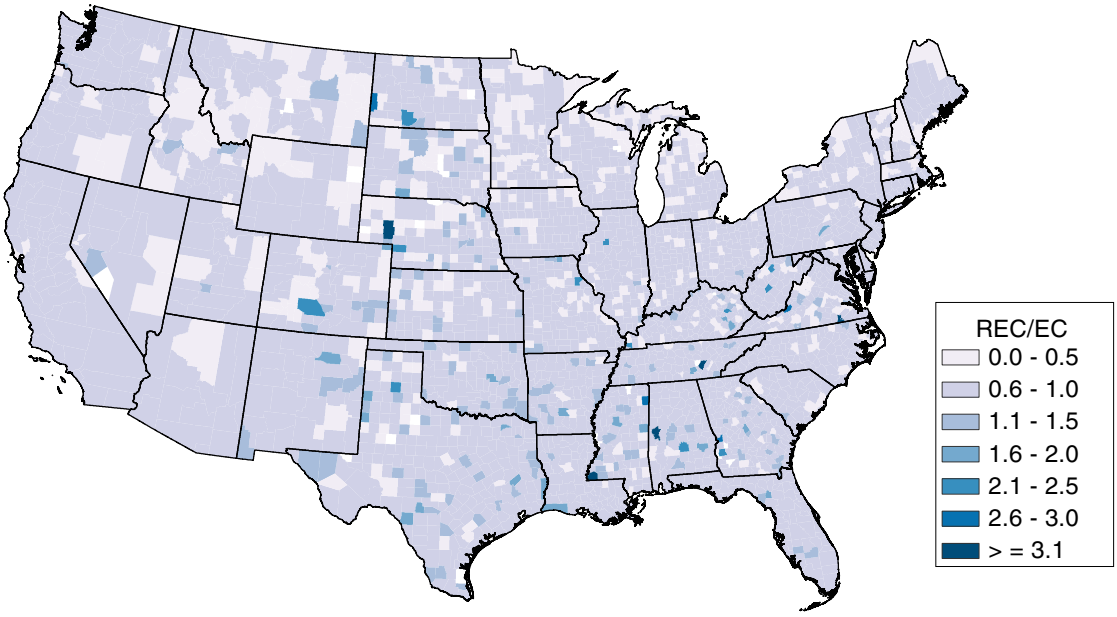
Difference between the number of FAFH establishments in NETS and the Economic Census, by county, 2012



Note: FAFH establishments in the Economic Census are in the following NAICS codes: 722320, 72251, 722511, 722513, 722514, 722515. FAFH establishments in the restricted NETS sample are in the following SIC codes: 5812 – Eating Places (except 58129908 – Dinner Theater; and 58129906 – Contract Food Services). The two SIC codes are excluded from this analysis because we could not find a unique NAICS match. Nonshaded counties have no establishments, and a grey county has a ratio of 1.
NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home. NAICS = North American Industrial Classification System. EC= Economic Census.
Source: USDA, Economic Research Service using data from NETS and the Economic Census.

Figure 15

Difference between the number of FAFH establishments in ReCount and the Economic Census, by county, 2012



Note: FAFH establishments in the Economic Census are in the following NAICS codes: 722320, 72251, 722511, 722513, 722514, 722515. FAFH establishments in the restricted NETS sample are in the following SIC codes: 5812 – Eating Places (except 58129908 – Dinner Theater; and 58129906 – Contract Food Services). The two SIC codes are excluded from this analysis because we could not find a unique NAICS match. Nonshaded counties have no establishments, and a grey county has a ratio of 1.

NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home. NAICS = North American Industrial Classification System. REC= ReCount. EC= Economic Census.

Source: USDA, Economic Research Service using data from NETS and the Economic Census.

Discussion and Conclusion

This study compares the portrait of the food environment captured by three commercial datasets (i.e., NETS, TDLinx, and NPD ReCount) purchased by ERS for economic research by assessing how well these datasets merge to one another. In addition, the study compares these datasets to the 2012 Economic Census. In general, findings suggest that the portrayal of the FAH environment is reasonably consistent between NETS and TDLinx, with exceptions discussed below. While county-level counts of FAFH establishments are fairly similar between NETS and ReCount, most establishments found in one dataset are not found in the other. Difficulties in capturing all establishments likely arise due to issues surrounding business classification and features of the respective markets. Differences in the quality of matching across States and counties are not substantial, however, are notable in some instances.

One of the difficulties in capturing the FAH retail environment has been the increasing prevalence of nontraditional retailers, such as dollar stores, drug stores, and supercenters; their share of total retail food expenditures increased from 13.7 percent in 2000 to 21.5 percent in 2011 (USDA, ERS, 2017). Supercenters are frequently the primary food store of U.S. households, further illustrating the importance of nontraditional retailers in the FAH retail environment (e.g., Ver Ploeg et al., 2015; Volpe et al., 2017). Nevertheless, some of these nontraditional retailers may choose not to sell food, particularly perishables, or they predominately sell nonfood products depending on the expected demand in the area.

The NAICS and SIC codes categorize supercenters under General Merchandise Stores (NAICS: 452, SIC: 53) because of the wide variety of products offered by these establishments. In NETS, supercenters were found in both the Warehouse Club and Discount Department categories alongside other stores that do not sell food, while TDLinx categorizes supercenters under the Grocery Trade Channel.[26] Thus, while some stores under Discount Department in NETS match to a Supercenter Subchannel establishment in TDLinx, the unmatched establishments in NETS could be found under the Category Killer Trade Channel—Discount Department Store Subchannel in TDLinx, which is not acquired by ERS because they are assumed to generally not be food purveyors. This discrepancy in classifications may also explain why a higher percentage of FAH establishments in TDLinx match to NETS rather than vice versa.

Another discrepancy in store classifications could arise from NETS providing detailed classifications for some stores but not others, particularly Gasoline Service Stations. While some stores are provided with a detailed SIC code for Gasoline Service Stations With Convenience Stores (SIC: 55419904), others are classified under the broader category Gasoline Service Stations (SIC: 55410000). In contrast, stores belonging to the TDLinx Convenience Store Trade Channel are further categorized into either the Gas Station/Kiosk or Conventional Subchannels according to store size and the variety of products offered.[27] Thus, the unmatched establishments could be found in Gas Station/Kiosk in TDLinx, which is not acquired by ERS.

[26] Warehouse Clubs and Supercenters in NAICS (452910) corresponds to Warehouse Club Stores in SIC (53999906), and Discount Department Stores in NAICS (452112) corresponds to Discount Department Stores in SIC (53119901).

[27] TDLinx also has a Subchannel Military Convenience Stores for convenience stores located within the confines of a military installation.

In addition to discrepancies in classification, the low percentage of NETS FAH establishments matching to TDLinx may also be due to NETS capturing very small FAH retailers that TDLinx excludes from the Grocery Trade Channel. TDLinx uses the Food Marketing Institute definition of a supermarket—"a full-line, self-service grocery store with an annual sales volume of $2 million or more" (Nielsen, 2010). Although TDLinx includes stores with $1 million to $2 million in annual sales under the Superette Subchannel, it does not classify smaller grocery stores with less than $1 million in annual sales (e.g., bodegas); these smaller grocery stores are in the Extended Master Channel. This could be part of the reason that Food Stores in NETS (SIC 54) only match to 19.1 percent of establishments in TDLinx. It may also be part of the reason that the number of establishments reported by TDLinx is more similar to the number in EC than to the number in NETS; the EC does not send its survey to most very small firms to reduce the burden on businesses.[28] However, the higher number of establishments reported by NETS on a county level, compared with the number in the EC, may result from our NETS sample including establishments that have an FAH SIC code in any of the six levels of SIC codes; the EC only provides the primary NAICS code.[29]

Comparing sales estimates using matched stores between TDLinx and NETS restricted by SIC codes indicates that for the majority of establishments, TDLinx reports a higher estimate of sales than NETS. This may be partially due to TDLinx providing the best estimate of monthly sales for each establishment, intended to be used as an indicator of store size rather than an actual retail sales report; the sales reported by TDLinx are not meant to be time trendable (Nielsen, 2010). In contrast, NETS is a time-series database where sales are calculated to be time trendable (Walls & Associates, 2013). Thus, the sales information provided in NETS may be closer to the actual sales for each establishment, while the TDLinx sales information may provide the expected sales given the store's characteristics. Nevertheless, because supercenters are categorized under the Grocery Channel, TDLinx reports supercenters' grocery sales rather than total store sales; NETS reports total store sales for supercenters.

In contrast to the relative success of matching TDLinx to NETS, only roughly a third of establishments in ReCount found matches in NETS. This may be driven by characteristics of the FAFH industry that make collecting data on FAFH establishments inherently more difficult than collecting data on FAH establishments. Most restaurants are small, independent operations less likely to be part of business registries such as InfoUSA or have a strong web presence than chain restaurants, which are key information sources in the development of ReCount. This idea is consistent with the finding of relatively higher match rates for QSRs, which tend to be chains. Furthermore, both datasets report fewer FAFH establishments than the EC in about 90 percent of the counties, which highlights the difficulty in capturing these small, independent restaurants.

While matching NETS to ReCount and vice versa produced a similar percentage of matches (roughly 33 to 36 percent), the percentage of matches is low, suggesting that the stores reported in each dataset are substantially different. Differences in data collection may drive this discrepancy, as may stylized features of the FAFH industry. For example, FAFH establishments experience high turnover rates, with as many as 26 percent of businesses failing in their first year of operation (Parsa et al., 2005). Therefore, if an establishment opens and closes within 1 year's time, one data source could report the establishment as open while the other never reports it.

[28] See Economic Census Frequently Asked Questions for additional details.

[29] NETS provides a crosswalk to convert its eight-digit SIC codes to NAICS codes, which we used to compare datasets.

Despite these difficulties in capturing the food environment, some groups of establishments have high match rates and are better represented by these various datasets. For instance, for FAH, we find that wholesale clubs in TDLinx have the highest percentage of matches in NETS (at 92.8 percent), followed by drug stores (at 85.6 percent). In turn, warehouse club stores and drug stores in NETS also have the highest percentage of matches in TDLinx (45.7 percent and about 43.0 percent, respectively).

Match percentages tend to be similar by county for both FAH and FAFH establishments, although urban counties see more matches than rural counties. For FAH, this characterization applies across classifications except for wholesale clubs, where the match rate is equal across rural and urban counties. For FAFH, there are higher match rates for QSRs in rural counties, likely reflecting the increased predominance of this format in recent years. Comparing low-poverty counties (i.e., those with a poverty rate of 20 percent or lower) to high-poverty counties reveals substantially similar match rates. Match rates for FAH and FAFH establishments are slightly higher in low-poverty counties, although the match rate was slightly higher in high-poverty counties for QSRs.

The main attraction of these three datasets is the store-level information provided by each. NETS provides time-trendable sales and employment information at the integer level, making it ideal for researchers and policymakers who wish to compare store sales or employment over time. It also provides NAICS and SIC codes for each establishment, making it easier to merge with other datasets. However, TDLinx provides more information about each establishment that is not available in NETS, such as the number of checkout registers, square footage of the selling area, and whether certain products are sold at each establishment (e.g., gas and liquor). In addition, the classification system in TDLinx may be more appealing for policymakers and researchers who are interested in examining nontraditional FAH establishments that sell food, particularly supercenters. Furthermore, because TDLinx classifies supercenters as grocery stores, their sales are calculated for the grocery store portion rather than for the entire store. However, one of the drawbacks is that TDLinx reports the estimated sales in brackets. Furthermore, any changes in the calculation of sales are not applied to previous years and are thus not time trendable. While users should avoid using TDLinx to examine changes among establishments' sales over time, it would be appropriate to compare sales within a given year.

One of the main advantages of NPD ReCount is its four-tiered categorization system. However, our findings on commercial data capturing the FAFH environment indicate the importance of finding innovative techniques to gather information on FAFH establishments, particularly in light of the steadily increasing share of food expenditures spent on FAFH over the last few decades (USDA, ERS, 2016). Although commercial data provide historical and detailed information on food retailers, online data platforms that source at least some information from users might be used to supplement these datasets. For instance, Glaeser et al. (2017) illustrate how online reviews submitted by users on Yelp can be used to predict changes in the number of establishments listed in the Census Bureau's County Business Patterns. A similar technique could potentially be used to better capture the FAFH environment and very small retailers; however, the properties of these data (e.g., how they are collected, coverage of the market) are understudied. For example, it is unclear whether using online data sources such as Yelp will improve the timeliness of data. On one hand, new businesses will be added to online databases (or closed businesses removed) on a rolling basis; thus, new data could be captured by researchers in more frequent intervals. On the other hand, if users fail to report changes in a timely manner and no other collection procedure accounting for this possibility is in place, then the commercial data sources detailed herein may prove more accurate. Similarly, it is unclear

whether online data sources will better capture smaller businesses not observed in one or more of the commercial and survey data sources, especially if such businesses are not frequently patronized or reviewed by users of online data platforms.

References

Alviola IV, P.A., R.M. Nayga, Jr., M.R. Thomsen, D. Danforth, and J. Smartt. 2014. "The Effect of Fast-Food Restaurants on Childhood Obesity: A School Level Analysis," *Economics & Human Biology* 12: 110-119.

Alwitt, L.F., and T.D. Donley. 1997. "Retail Stores in Poor Urban Neighborhoods," *Journal of Consumer Affairs* 31(1): 139-164.

Barnatchez, K., L. Crane, and R. Decker. 2017. "An Assessment of the National Establishment Time Series (NETS) Database," Finance and Economics Discussion Series 2017-110. Washington: Board of Governors of the Federal Reserve System.

Binkley, J.K. 2008. "Calorie and Gram Differences Between Meals at Fast Food and Table Service Restaurants," *Applied Economic Perspectives and Policy* 30(4): 750-763.

Cadieux, R., T. Watson, and D. Bretheim. 2014. "Matching Rules: Too Loose, Too Tight, or Just Right?" SAS Global Forum Paper 1674-2014. Accessed from SAS website.

Caspi, C.E., G. Sorensen, S. Subramanian, and I. Kawachi. 2012. "The Local Food Environment and Diet: A Systematic Review," *Health & Place* 18(5): 1172-1187.

Courtemanche, C., and A. Carden. 2011. "Supersizing Supercenters? The Impact of Walmart Supercenters on Body Mass Index and Obesity," *Journal of Urban Economics* 69(2): 165-181.

Currie, J., S. DellaVigna, E. Moretti, and V. Pathania. 2010. "The Effect of Fast Food Restaurants on Obesity and Weight Gain," *American Economic Journal: Economic Policy* 2(3): 32-63.

Dunn, R.A., J.R. Sharkey, and S. Horel. 2012. "The Effect of Fast-Food Availability on Fast-Food Consumption and Obesity Among Rural Residents: An Analysis by Race/Ethnicity," *Economics & Human Biology* 10(1): 1-13.

Dunn, T. 2014. "Getting the Warm and Fuzzy Feeling With Inexact Matching," SAS Global Forum Paper 1316-2014. Accessed from SAS website.

Gebreab, S.Y., D.A. Hickson, M. Sims, S.B. Wyatt, S.K. Davis, A. Correa et al. 2017. "Neighborhood Social and Physical Environments and Type 2 Diabetes Mellitus in African Americans: The Jackson Heart Study," *Health & Place* 43: 128-137.

Glaeser, E., H. Kim, and M. Luca. 2017. "Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity," National Bureau of Economic Research Working Paper 24010.

Guthrie, J., B. Lin, and E. Frazao. 2002. "Role of Food Prepared Away From Home in the American Diet, 1977-78 Versus 1994-96: Changes and Consequences," *Journal of Nutrition Education and Behavior* 34 (3): 140-150.

Kaufman, P.R., C. Dicken, and R. Williams. 2014. *Measuring Access to Healthful, Affordable Food in American Indian and Alaska Native Tribal Areas*, Economic Information Bulletin No. 131, U.S. Department of Agriculture, Economic Research Service.

Kunkle, G. 2011. "Business Establishment Employment Data: Nets Versus Es-202," Edward Lowe Foundation's Institute for Exceptional Growth Companies.

Kyureghian, G., R.M. Nayga Jr., and S. Bhattacharya. 2012. "The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effects Analysis," *Applied Economic Perspectives and Policy* 35(1): 69-88.

Levin, D., D. Noriega, C. Dicken, A. Okrent, M. Harding, and M. Lovenheim. 2018. *Examining Food Store Scanner Data: A Comparison of the IRI InfoScan Data With Other Datasets*, 2008-2012, Technical Bulletin No. 1949, U.S. Department of Agriculture, Economic Research Service.

Lin, B.-H., and J. Guthrie. 2012. *Nutritional Quality of Food Prepared at Home and Away From Home, 1977-2008*, Economic Information Bulletin No. 105, U.S. Department of Agriculture, Economic Research Service.

Ma, X., S.E. Battersby, B.A. Bell, J.D. Hibbert, T.L. Barnes, and A.D. Liese. 2013. "Variation in Low Food Access Areas Due to Data Source Inaccuracies," *Applied Geography* 45: 131-137.

McLaughlin, P.W., and C. Dicken. 2018. "Evolution of the Food-Away-From-Home Industry: Recent and Emerging Trends," Chapter 6 in *America's Eating Habits: Food Away From Home*, Economic Information Bulletin No. 196, U.S. Department of Agriculture, Economic Research Service.

Moore, L.V., and A.V. Diez Roux. 2006. "Associations of Neighborhood Characteristics With the Location and Type of Food Stores," *American Journal of Public Health* 96(2): 325-331.

Neumark, D., J. Zhang, B. Wall. 2005. "Employment Dynamics and Business Relocation: New Evidence From the National Establishment Time Series," National Bureau of Economic Research Working Paper 11647.

Nielsen. 2010. "TDLinx Data Dictionary," internal document provided to USDA, ERS.

NPD. 2015. "Fall 2015 Release Notes," internal document provided to USDA, ERS.

Parsa, H., J. Self, D. Njite, and T. King. 2005. "Why Restaurants Fail," *Cornell Hotel and Restaurant Administration Quarterly* 46 (3): 304-322.

Pearson, T., J. Russell, M.J. Campbell, and M.E. Barker. 2005. "Do 'Food Deserts' Influence Fruit and Vegetable Consumption?—A Cross-Sectional Study," *Appetite* 45(2): 195-197.

Rahkovsky, I., Y. Jo, and L. Mancino. 2018. "How Food Environment and Proximity to Restaurants Affect Nutritional Quality," Chapter 8 in *America's Eating Habits: Food Away From Home*, Economic Information Bulletin No. 196, U.S. Department of Agriculture, Economic Research Service.

Rahkovsky, I., and S. Snyder. 2015. *Food Choices and Store Proximity*, Economic Information Bulletin No. 195, U.S. Department of Agriculture, Economic Research Service.

Rhone, A., M. Ver Ploeg, C. Dicken, R. Williams, and V. Breneman. 2017. *Low-Income and Low-Supermarket-Access Census Tracts, 2010-2015*, Economic Information Bulletin No. 165, U.S. Department of Agriculture, Economic Research Service.

Rummo, P.E., P. Gordon-Larsen, and S.S. Albrecht. 2015. "Field Validation of Food Outlet Databases: The Latino Food Environment in North Carolina, USA," *Public Health Nutrition* 18(6): 977-982.

Sweitzer, M., D. Brown, S. Karns, M. Muth, P. Siegel, and C. Zhen. 2017. *Food-at-Home Expenditures: Comparing Commercial Household Scanner Data From IRI and Government Survey Data*, Technical Bulletin No. 1946, U.S. Department of Agriculture, Economic Research Service.

Thomsen, M.R., R.M. Nayga, Jr., P.A. Alviola IV, and H.L. Rouse. 2015. "The Effect of Food Deserts on the Body Mass Index of Elementary Schoolchildren," *American Journal of Agricultural Economics* 98(1): 1-18.

Todd, J. 2017. "Changes in Consumption of Food Away From Home and Intakes of Energy and Other Nutrients Among US Working-Age Adults, 2005-2014," *Public Health Nutrition*, 20 (18): 3238-3246.

U.S. Department of Agriculture, Economic Research Service (ERS). "Retail Trends." Last updated December 7, 2017. Accessed from the Food Markets & Prices topic page on the ERS website in March 2018.

U.S. Department of Agriculture, Economic Research Service (ERS). "Trends in U.S. Food Expenditures." Last updated October 20, 2016. Accessed from the Food Expenditures data product on the ERS website in April 2018.

Ver Ploeg, M., L. Mancino, J. Todd, D. Clay, and B. Scharadin. 2015. *Where Do Americans Usually Shop for Food and How Do They Travel To Get There? Initial Findings From the National Household Food Acquisition and Purchase Survey*, Economic Information Bulletin No. 138, U.S. Department of Agriculture, Economic Research Service.

Ver Ploeg, M., V. Breneman, T. Farrigan, K. Hamrick, D. Hopkins, P. Kaufman, B.-H. Lin, M. Nord, T. Smith, R. Williams, K. Kinnison, C. Olander, A. Singh, and E. Tuckermanty. 2009. *Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences: Report to Congress*, Administrative Publication No. 036, U.S. Department of Agriculture, Economic Research Service.

Volpe, R., and A.M. Okrent. 2012. *Assessing the Healthfulness of Consumers' Grocery Purchases*, Economic Information Bulletin No. 102, U.S. Department of Agriculture, Economic Research Service.

Volpe, R., A. Kuhns, and E. Jaenicke. 2017. *Store Formats and Patterns in Household Grocery Purchases*, Economic Information Bulletin No. 167, U.S. Department of Agriculture, Economic Research Service.

Volpe, R., A. Okrent, and E. Leibtag. 2013. "The Effect of Supercenter-Format Stores on the Healthfulness of Consumers' Grocery Purchases," *American Journal of Agricultural Economics* 95(3): 568-589.

Walker, R., C. Keane, and J. Burke. 2010. "Disparities and Access to Healthy Food in the United States: A Review of Food Deserts Literature," *Health and Place* 16: 876-884.

Walls & Associates. 2013. “National Establishment Time-Series (NETS) Database: 2013 Database Description.” Internal document provided to USDA, ERS.

Zenk, S.N., A.J. Schulz, B.A. Israel, S.A. James, S. Bao, and M.L. Wilson. 2005. "Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit," *American Journal of Public Health* 95(4): 660-667.

Appendix

Appendix table 1

Share of TDLinx-NETS matches, full dataset and restricted by SIC codes, in three most frequent NETS classifications

TDLinx primary classification NETS classification	Full matches	TDLinx primary classification NETS classification	Restricted matches*
	Percent		*Percent*
Convenience Store	90.1	Convenience Store	96.8
Grocery Stores	63.9	Grocery Stores	78.2
Gasoline Service Stations	23	Gasoline Service Stations	17.9
Retail Stores, NEC	3.2	Misc. General Merchandise Stores	0.7
Grocery	88.5	Grocery	94.1
Grocery Stores	78.4	Grocery Stores	82.8
Department Stores, Discount	7.6	Department Stores, Discount	8.7
Miscellaneous Food Stores	2.5	Miscellaneous Food Stores	2.6
Drug	97.6	Drug	99.6
Drug Stores and Proprietary Stores	95.9	Drug Stores and Proprietary Stores	99.2
Drugs, Proprietaries, and Sundries	1	Grocery Stores	0.2
Retail Stores, NEC	0.7	Variety Stores	0.1
Mass Merchandiser	97.1	Mass Merchandiser	99.4
Variety Stores	80.5	Variety Stores	86.3
Department Stores, Discount	13.1	Department Stores, Discount	12.7
Retail Stores, NEC	3.5	Misc. General Merchandise Stores	0.4
Wholesale Club	96.1	Wholesale Club	99.4
Misc. General Merchandise Stores	91.1	Misc. General Merchandise Stores	96.6
Retail Stores, NEC	2.9	Drug Stores and Proprietary Stores	2.1
Drug Stores and Proprietary Stores	2.1	Department Stores, Discount	0.8

Note: * indicates restricted NETS sample, which includes establishments in the following SIC codes: 54 – Food Stores; 53119901 – Department Stores, Discount; 53310000 – Variety Stores; 53999903 – Country General Stores; 53999906 – Warehouse Club Stores; 55410000 – Gasoline Service Stations; 55419904 – Gasoline Service Stations With Convenience Stores; and 5912 – Drug Stores and Proprietary Stores. NEC = not elsewhere classified. SIC = Standard Industrial Classification. NETS = National Establishment Time Series database.

Source: USDA, Economic Research Service using data from TDLinx and NETS.

Appendix table 2

Share of ReCount-NETS matches, full dataset and restricted by SIC codes, by top three NETS FAFH classifications

Industry group NETS classification	Full matches	Industry group	Restricted matches*
Full-service restaurants			
	Percent		*Percent*
Casual Dining	92.5	Casual Dining	99.4
Eating and Drinking Places	89.9	Eating and Drinking Places	99
Grocery Stores	1.4	Hotels and Motels	0.2
Groceries and Related Products	1.2	Miscellaneous Personal Services	0.2
Family Dining	93	Family Dining	99.2
Eating and Drinking Places	89.4	Eating and Drinking Places	98.7
Gasoline Service Stations	2	Gasoline Service Stations	0.3
Grocery Stores	1.6	Grocery Stores	0.2
Fine Dining	92.6	Fine Dining	98.7
Eating and Drinking Places	88	Eating and Drinking Places	96.1
Hotels and Motels	3.2	Hotels and Motels	1.3
Management and Public Relations	1.4	Gasoline Service Stations	1.3
Unclassified	83.1	Unclassified	99.1
Eating and Drinking Places	78	Eating and Drinking Places	98.7
Grocery Stores	3.2	Hotels and Motels	0.2
Management and Public Relations	1.8	Grocery Stores	0.2
Quick-service restaurants			
Sandwich/Mexican	95.7	Sandwich/Mexican	99.5
Eating and Drinking Places	89.2	Eating and Drinking Places	98.9
Grocery Stores	4.5	Grocery Stores	0.5
Groceries and Related Products	2	Dairy Products Stores	0.1
Snack	80.9	Snack	98.4
Eating and Drinking Places	46.6	Eating and Drinking Places	93.8
Retail Bakeries	25.6	Retail Bakeries	3.7
Miscellaneous Food Stores	8.8	Miscellaneous Food Stores	0.8
Specialty	96.6	Specialty	99.8
Eating and Drinking Places	95.4	Eating and Drinking Places	99.6
Grocery Stores	0.9	Grocery Stores	0.1
Miscellaneous Shopping Goods Stores	0.4	Miscellaneous Personal Services	0

Note: * indicates restricted NETS sample, which includes establishments in the following SIC code: 5812 – Eating Places. NETS = National Establishment Time Series database. SIC = Standard Industrial Classification. FAFH = food away from home.

Source: USDA, Economic Research Service using data from ReCount and NETS.