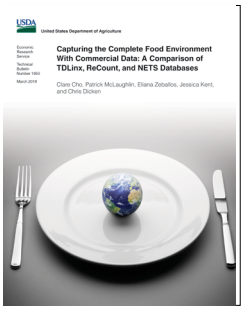
United States Department of Agriculture

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.