



United States Department of Agriculture

Using Proprietary Household and Retail Scanner Data in Food Policy Research

Jay Variyam

Megan Sweitzer

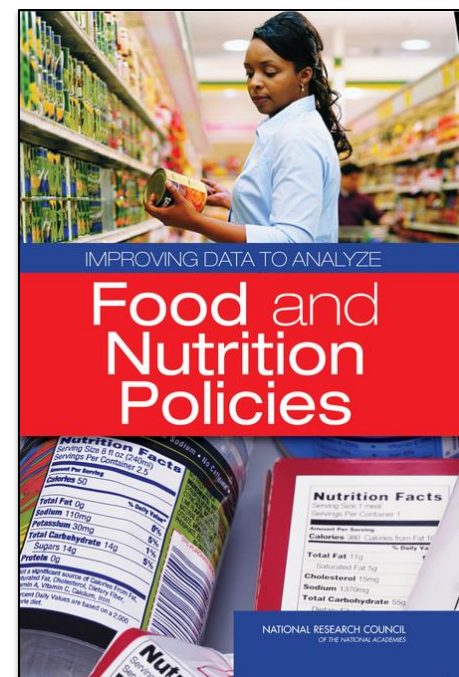
USDA – Economic Research Service

December 16, 2015



Background and Motivation

- ERS requested CNSTAT panel advice for improving data infrastructure on food consumption and nutrition
- Recommendations in 2005 book *Improving Data to Analyze Food and Nutrition Policies*:
 - Continue to explore the use of proprietary retail and household scanner data
 - Examine the quality, characteristics, and representativeness of the data



Background and Motivation

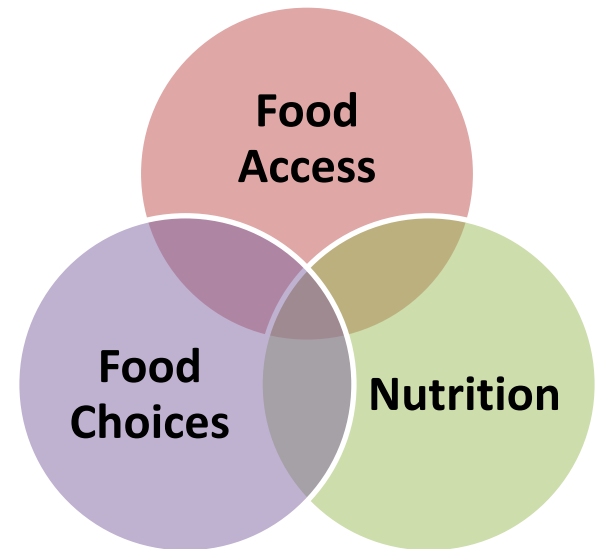
- Consumer Data and Information Program
- Targeting new data investments to fill critical gaps in USDA's food consumption data
- Importance of understanding what people eat
 - Consumption and expenditure patterns
 - Consumer responses to price changes, health concerns, new products
 - Relationships between food choices, diet, and health outcomes, with emphasis on low-income populations



The Mission of The Consumer and Food Data System

Develop data and information that support food economics research, including

- The determinants of food access, consumer food choices, and
- The nutrition and health outcomes



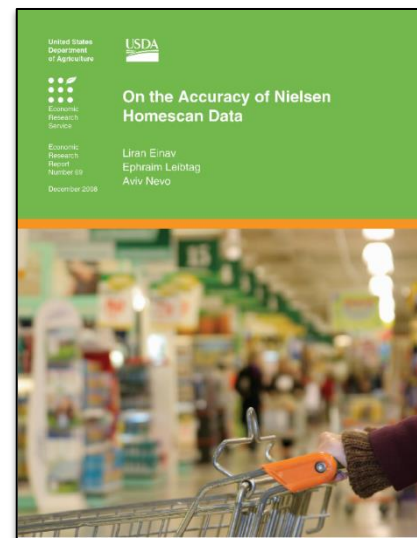
Scanner Data at ERS

- Nielsen Homescan
 - Initial data investment: 1998-2010
 - Household panel data
 - Used in 135 research products or program and regulatory applications
- IRI Consumer Network & InfoScan
 - Expanded set of data: 2008-present
 - Household panel, retail point-of-sale, health information, and nutrition data



Evaluating Data Quality

- *On the Accuracy of Nielsen Homescan Data*
 - Validation study of Homescan using itemized retailer data
 - Detailed record-matching and analysis to find:
 - What level of measurement error exists
 - Does it matter for analyses
 - Reporting errors are in line with other research data sets
 - Included two recommendations, one of which was implemented in future data acquisitions



Einav, Liran, Ephraim Leibtag, and Aviv Nevo. *On the Accuracy of Nielsen Homescan Data*. ERR-69, U.S. Dept. of Agriculture, Econ. Res. Serv. December 2008.



Evaluating Data Quality

- *Understanding Differences in Self-Reported Expenditures between Household Scanner Data and Diary Survey Data: A Comparison of Homescan and Consumer Expenditure Survey*
 - Matched surveys by food product category
 - Found substantial differences between Homescan and CES expenditures
 - Explained that many differences in expenditures can be explained by household demographics
 - Discussed implications for analysis

Zhen, C., J. L. Taylor, M. K. Muth, and E. Leibtag. "Understanding Differences in Self-Reported Expenditures between Household Scanner Data and Diary Survey Data: A Comparison of Homescan and Consumer Expenditure Survey," *Review of Agricultural Economics* 31 (3): 470-92, Fall 2009.



IRI Data Description

- Household-based scanner data
 - Itemized food-at-home purchases for 120,000 households
 - Household demographics
 - Supplementary health information and prescription drug surveys
- Retail point-of-sale data
 - Quantity and dollar sales by UPC by store by week
 - About 6.5 billion transaction records per year
- Product dictionaries
 - Product information for one million products, including nutrition data
- Retail store information



Evaluating Data Quality

- Basic data quality and process improvement
 - Required non-trivial level of data infrastructure
 - Developed set of quality checks
 - Worked with vendor to implement in delivery process
- Data modifications for researcher-friendly formatting



Evaluating Data Quality

- IRI statistical properties studies
 - Methodology report
 - Four comparison studies:
 - Household expenditures to CES
 - Retail sales to TDLinx, NETS, CBP, EC
 - Medical information to NHIS, MEPS
 - Nutrition data to Gladson
- Upcoming workshop on applying IRI data in food policy studies



Combining IRI with Other Data Sources

- FoodAPS
 - Streamlined process for data access
- USDA Nutrient Databases
 - Creating a tool to estimate prices for foods as consumed
 - Automating two ERS data products
 - Fruit and Vegetable Prices
 - Quarterly Food-at-Home Price Database
- Geospatial and store characteristics data
 - Linking retail sales with more complete picture of food environment





United States Department of Agriculture

Jay Variyam
Division Director – Food Economics Division
jvariya@ers.usda.gov

Megan Sweitzer
megan.sweitzer@ers.usda.gov

