Sources and Analyses of Food Label Data from Proprietary Sources

Workshop on Food Prices and Forecasting

New Orleans, July 31-August 1, 2024

Mary K. Muth

RTI INTERNATIONAL

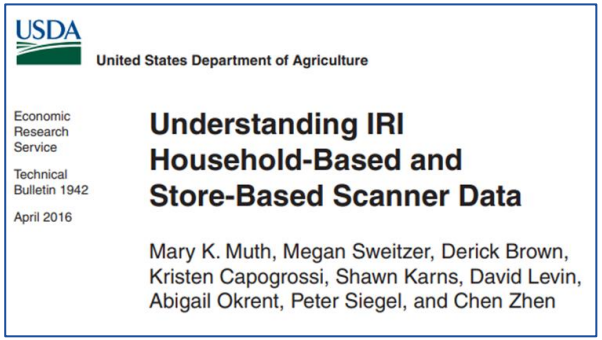
Overview

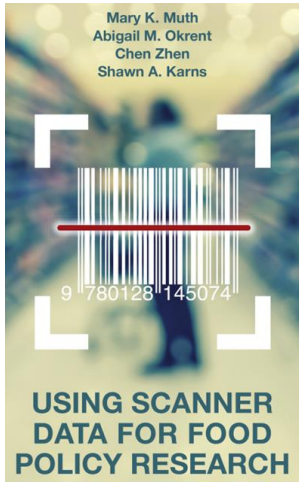
- **Store and household scanner data include product attributes at the barcode level that can be used to classify products for conducting analyses of food prices at a disaggregated level.**
- Content of food label data
  - Nutritional or ingredient content
  - Other attributes
- Sources of data
- Example applications
  - Development of the product code mapping underlying the F-MAP
  - Hedonic analyses of product attributes
- Challenges in using the data

Content of Food Label Data: Nutrition, Claims, Other Attributes

- Nutritional content
  - Serving size and no. of servings
  - Calories
  - Numeric values for each nutrient on the Nutrition Facts label
  - % Daily value for each nutrient
- Claims
  - Nutrient content (more common)
  - Health claims (less common)
- Other potential attributes
  - Ingredients
  - Organic content, natural, kosher, bioengineered free, gluten, etc.

- Resources
  - ERS Technical Bulletin 1942 (2016)
  - Using Scanner Data for Food Policy Research, Muth et al. (2020)
  - Sample data on Nutritionix website: https://www.nutritionix.com/database





Providers of Food Label Data at the Barcode Level

Provider	Countries
Circana	US
FoodSwitch	Australia, Asian countries, US, UK
Kantar	European countries
Label Insight (NielsenIQ)	US
Mintel	> 60 countries
Nutritionix	US
Syndigo (Gladson)	US, Canada
USDA Branded Products Database	US

- Data fields may be
  - Provided with scanner data
  - Appended to scanner data by barcode
- Data sources
  - Coded from images taken by field data collectors in stores, provided by manufacturers, uploaded into fitness apps, or obtained on websites
  - Datasets provided by manufacturers

Example Application: F-MAP

Classify ~600K barcodes to 90 ERS Food Purchase Groups (EFPGs) to calculate price measures

- Prepackaged products with barcodes (>90%)
- Perishable products with price lookup codes or retailer-specific barcodes (<10%)

Programming code logically assigns using multiple data fields

E.g., category, product type, product description, “type of” fields, fat content, flavor/scent, style, grain claim

Example Application: Hedonic Analyses

- Hedonic analyses allow estimation of implicit prices of product attributes (Rosen, 1974)
  - Using estimated coefficients, can calculate percentage effect of each label attribute using Kennedy's (1981) adjustment
  - Numerous applications using scanner data are in the literature

- Semi-log regression from Giombi, Muth, & Levin (2018) for soup products
  - Ln P = f(label statements, nutrient values, branded/PL, package size)
  - Price calculated as total dollars divided by total units sold for each barcode and then dividing by number of ounces (on an as-consumed basis)
  - Compared nutrient values measured in grams versus % Daily Value
  - Results show estimated effects of attributes on implicit price per ounce
    - E.g., organic (+), saturated fat content (-), private label (-), and package size (-)

Challenges in Using Food Label Data

Many barcodes do not have matching label data

Products with a higher sales volume are more likely to have label data

Using multiple data sources can increase match rate

Must update data for products each year

About 15-20% of barcodes are new each year but could represent a change to an existing product

Must calculate sales volumes to determine if barcode is active

Records for discontinued products are not necessarily purged

When using nutrient data, need to understand limitations resulting from regulatory specifications

Rounding of nutrients and %daily values (e.g., <0.5g per serving may show as 0)

Some products have nutrients per serving and per package

Need to ensure using comparable fields across products

Ingredient lists are typically text strings

Must parse text strings to use ingredients but often inconsistent terminology

Thank you

Contact: Mary K. Muth| email: muth@rti.org