

Preparation of the 1999 ACNielsen Homescan Panel Data

The data set used in this study is the 1999 ACNielsen Homescan panel data, purchased by USDA in 2002 for use in several projects supported by Economic Research Service. This project addressed the goal of determining the types of nonalcoholic beverages purchased by U.S. households in different demographic segments. One of the benefits of using these data for this study was to determine if this type of scanner data could be used as a viable alternative to other more costly and scarce data sources.

Data Description

The ACNielsen Homescan data set is drawn from a sample of households that are demographically balanced within 19 markets and 4 Census regions in the United States. Sample households are projected to market universes and weighted so that a nationally representative sample captures buying patterns of U.S. consumers. The ACNielsen Homescan data are unique in that the information is similar to a survey. Each household was supplied with a scanner device that was used to record all items purchased at different retail trade locations throughout a given time period. Each panelist represented a unique household, with each household having 17 known demographic characteristics (table 5).

The ACNielsen households surveyed represented 52 different cities, which was 84.34 percent of the surveyed respondents, as well as unidentified rural areas, which represented 15.66 percent of the surveyed households (table 6).

Table 5—Demographic characteristics of panelist households

Panelist descriptives	Number of categories within each description
1. Household size	9
2. Household income	16
3. Age of female head ¹	10
4. Age of male head ²	10
5. Age and presence of children	8
6. Male head employment ²	5
7. Female head employment ¹	5
8. Male head education ²	7
9. Female head education ¹	7
10. Marital status	5
11. Male head occupation ²	12
12. Female head occupation ¹	12
13. Household composition	8
14. Race	4
15. Hispanic origin	2
16. Region	4
17. Scantrack market identifier	53

¹ Female head of household is the primary person making food purchase decisions.

² Male head of household is the primary person making food purchase decisions.

Source: ERS analysis of ACNielsen Homescan panel data.

The survey covered 4 regions of the lower 48 States of the United States—East, Central, South, and West.

The regional representation of the surveyed U.S. households was similar to the 1999 Bureau of Census regional representation by percent of households surveyed:

- East—20.0 percent (Census); 20.3 percent (ACNielsen)
- Central—24.0 percent (Census); 25.3 percent (ACNielsen)
- South—34.0 percent (Census); 34.3 percent (ACNielsen)
- West—22.0 percent (Census); 20.0 percent (ACNielsen).

Although ACNielsen Homescan data include purchases of all consumer items bought during a specified time period, the nationally representative data that was used in this study included only consumer purchases of food items. Household level purchase data and demographic information were included for 7,195 household panelists who were in the sample during at least 10 out of 12 months beginning January 3, 1999 through January 1, 2000. The household level food purchase data are divided into four product type groups:

- (1) Dry grocery (4,111,719 records)
- (2) Dairy (873,899 records)
- (3) Frozen (1,002,851 records)
- (4) Random weights (507,306 records).

Table 6—Panelist households' locations

Scantrack market	Percent of households surveyed	Scantrack market	Percent of households surveyed
Rural	15.66	San Diego	0.61
Boston	1.30	St. Louis	0.96
Chicago	10.46	Tampa	0.77
Houston	0.56	Baltimore	4.30
Indianapolis	1.27	Birmingham	0.25
Jacksonville	0.28	Buffalo-Rochester	1.04
Kansas City	0.76	Hartford-New Haven	1.17
Los Angeles	11.26	Little Rock	0.15
Suburban New York	5.47	Memphis	0.08
Urban New York	3.81	New Orleans-Mobile	0.18
Non-urban New York	2.79	Oklahoma City-Tulsa	0.13
Orlando	0.48	Phoenix	1.83
San Francisco	0.64	Raleigh-Durham	0.23
Seattle	0.71	Salt Lake City	1.57
Atlanta	13.79	Columbus	0.58
Cincinnati	0.94	Washington, DC	8.83
Cleveland	1.01	Albany	0.49
Dallas	0.40	Charlotte	0.56
Denver	0.86	Des Moines	0.49
Detroit	1.32	Grand Rapids	0.91
Miami	0.64	Louisville	0.18
Milwaukee	0.63	Omaha	0.56
Minneapolis	0.56	Richmond	0.28
Nashville	0.16	Sacramento	0.48
Philadelphia	1.80	San Antonio	7.51
Pittsburgh	1.43	Syracuse	1.45
Portland, Oregon	1.09		

Source: ACNielsen Homescan data.

Each of the four groups contained numerous product modules, with each product module further subdivided into brand, size, flavor, form, formula, container, style, type and variety. Each product was represented by a unique unit product code (UPC). In addition to demographic information, total expenditure and quantity information were recorded for each transaction. This information was used to impute the price per unit for the food items analyzed in this study.

Data Selection Process

This step included the process of selecting and organizing the data so that it was usable for analytical and descriptive purposes. The primary objective was to understand consumer demand and nutritional issues associated with nonalcoholic beverages purchased for at-home use. The beverages included in this study were all milk items, isotonics (sports drinks), bottled water, fruit juices and drinks, coffee and tea, and carbonated and noncarbonated soft drinks. (See box, “Nonalcoholic beverage categories.”)

The process of obtaining a usable data set included determining which product modules were needed to construct the appropriate final data set. Of the many hundreds of modules in the data set, 53 beverage modules were

Nonalcoholic beverage categories

Ready-To-Drink (RTD) fruit juices not frozen
Apple juice not frozen
Orange juice not frozen
Other fruit juices not frozen
Ready-To-Drink fruit drinks
Isotonics–sports drinks
Powdered soft drinks
Vegetable juices and drinks
Tea
Tea–regular
Tea–decaffeinated
Coffee
Coffee–regular
Coffee–decaffeinated
Carbonated soft drinks
Carbonated soft drinks–regular
Carbonated soft drinks–low calorie
Bottled water
Milk–flavored and unflavored
Flavored milk
Unflavored milk
Flavored milk–low fat
Flavored milk–whole
Unflavored milk–whole
Unflavored milk–2%
Unflavored milk–1%
Unflavored milk–skim
Fruit juices frozen
Frozen fruit drinks
Other fruit juices frozen
Apple juice frozen

Source: ACNielsen Homescan data.

selected for analysis. Several of the 53 modules were further disaggregated or aggregated to create other modules, which also were used in constructing the final data set. The final data set contained 77 different beverage product modules (data table 13). The purpose of the aggregation/disaggregation process was to group the beverages in modules that would allow for a thorough analysis. Not only might the effects of the particular beverage, such as milk, be important, but the effects from different varieties of milk—flavored, unflavored skim, low fat, etc.—also might be important. A listing of the different aggregations of modules is included in data table 14.

Each of the 77 beverage modules was converted into the common measure of gallons in order to have valid comparisons with the other modules. This process required two things: First, a knowledge of the form, size and quantity of the products in the modules, and second, the rate of conversion for each form, size, and quantity. The first step was simple since the product form, size, and quantity were included in each record. The second step was not as simple and in some cases required actual physical examination of a product before it could be converted into gallons.

After the product modules were extracted, created, and converted to gallons, further checking of the actual data was necessary. A very limited number of records were unusable because expenditures were missing. After removing the records that were unusable, the imputation of prices (ratio of expenditures to quantities) for the remaining records were completed. Price outliers, defined as imputed prices greater than five standard deviations from their means, were flagged and omitted. Records corresponding to these price outliers also were eliminated from the analysis. In data table 14, we delineate the number of missing expenditure records by product module together with the number of records corresponding to price outliers by product module.

The 1999 ACNielsen Homescan data include transactions that were made during the year and recorded by a scanner at home, and as such could be considered a panel data set with both cross-sectional and time-series characteristics. However, the sporadic nature of the time-series entries associated with the data set make it more practical to convert the home scan data to a cross-sectional type of data set. Since each record included purchase transactions by a particular household, each recorded transaction in the same product module was identified and combined to create an annual household consumption (purchase) quantity and expenditure amount for each module. Then, the household total purchase quantity and total expenditure amounts were used to impute an average annual household price for each module. In this study we explicitly assume that all beverage items purchased are consumed by the household. Thus we equate household purchases with household consumption and intake. For a summary of the 77 annual consumption figures, expenditure figures, and average annual prices, see data table 13. These descriptive statistics take into account the projection factors used to make this sample of households nationally representative.

This report serves to summarize the work done on the ERS cooperative project, “Demand Projections Segmented by Income for the Highly Competitive Nonalcoholic Beverage Complex Using the ACNielsen Homescan Panel Data.” As the data are for 1999, this work constitutes a baseline analysis of economic and nutritional issues in conjunction with a choice of

77 nonalcoholic beverages, for both aggregate and disaggregated analysis. This 1999 baseline will be useful in future work for evaluating consumer impacts of the advice issued in 2000 and 2005 by the Dietary Guidelines for Americans, which gave increased emphasis to beverage choice, particularly soft drinks.

Cross-Tabulations of Household Purchases (Consumption)

Cross-tabulations were used to examine the relationship of household purchases of nonalcoholic beverages with various demographic factors. With this procedure, a specific demographic variable is identified and summary statistics are computed for the records in the data set that correspond to only those demographic criteria. For example, the average consumption in gallons per household of a selected beverage is calculated for each demographic category. It should be noted that the calculated averages include only the households that purchased the selected beverage. After all demographics variables are tabulated, comparisons can be made. To illustrate, the demographic variable region includes four categories: East, Central, South, and West. Average levels of consumption for the households in each region were calculated. A comparison among the households in the four regions quickly reveals if there is a difference in the level of purchases from one region to another. The number of households purchasing each beverage in each demographic category also is included in this treatment.

The demographic variables used in the analysis include poverty status, household size, age of female head, employment of female head, education of female head, race, region, Hispanic origin (ethnicity), and seasonality (data tables 15a and 15b). The beverage groupings to be analyzed in the cross-tabulations are shown in the box on page 21. To conform to space limitations, both aggregate groupings and disaggregated groupings of beverages are used rather than all 77 beverage products previously discussed.

A Look at Prices, Gallons, and Expenditures

The annual consumption (in gallons), expenditure (in dollars), and prices (dollars per gallon) for households who actually bought nonalcoholic beverages in 1999 are shown in data table 13. The statistics include a count of the number of households who purchased the nonalcoholic beverage in 1999, as well as the mean, median, standard deviation, minimum and maximum associated with gallons bought, prices paid, and expenditures made on nonalcoholic beverages.

To illustrate, 4,898 households bought bottled water; 5,304 bought tea; 5,584 bought coffee; 7,036 bought milk; and 7,041 bought carbonated soft drinks. These figures correspond to market penetration (the percent of respondents who actually consumed the beverage) of 68.1 percent for bottled water; 73.7 percent for tea; 77.6 percent for coffee; 97.8 percent for milk; and 97.9 percent for carbonated soft drinks. Carbonated beverages were the most popular beverage purchase by households and packaged tea was the least purchased item of the 77 product modules analyzed.

Average prices paid per gallon for bottled water were \$1.99; tea \$1.89; coffee, \$1.38; milk, \$3.08; and carbonated soft drinks, \$2.45. The least expensive beverages for 1999 are powdered soft drinks, coffee, and tea; the most expensive nonalcoholic beverages are ready-to-drink fruit juices, vegetable juices, and isotonic.

Average annual quantities and expenditures for households who bought various nonalcoholic beverages are as follows: ready-to-drink fruit juices, 13.47 gallons (\$60.35); ready-to-drink fruit drinks, 8.14 gallons (\$27.29); isotonic, 3.58 gallons (\$15.36); powdered soft drinks, 17.89 gallons (\$14.14); vegetable juices and drinks, 2.29 gallons (\$12.96); tea, 15.00 gallons (\$18.58); coffee, 43.06 gallons (\$42.81); carbonated soft drinks, 51.87 gallons (\$121.19); bottled water, 14.32 gallons (\$17.73); flavored milk, 2.32 gallons (\$9.80); unflavored milk, 33.32 gallons (\$90.78); frozen fruit juices, 6.77 gallons (\$20.77); and frozen fruit drinks, 3.61 gallons (\$9.63). The volume leaders in 1999 were carbonated soft drinks, coffee, and unflavored milk, in that order. Average expenditures are greatest for carbonated soft drinks, milk, ready-to-drink fruit juices, and coffee, in that order.

Poverty Status

Instead of using only the income demographic given in the ACNielsen Homescan data, a poverty threshold demographic also was calculated according to U.S. Census Bureau poverty specifications. Both income and household size were used for determining households below and above the poverty threshold. We are using 130 percent of poverty in this study because it is the cut-off level for food stamp eligibility and for free school meals. Analysis of the household income levels found that 423 of the 7,195 households fell into the below 130 percent of poverty range. The households above 130 percent of poverty purchased more orange juice, both frozen and not frozen (fig. 33, data table 16). Households below 130 percent of poverty purchased over 3 more gallons of powdered soft drinks a year and purchased over 7 more gallons of regular carbonated soft drinks per year when compared with households with incomes over 130 percent of poverty. Above 130 percent poverty households purchased more of the lower calorie soft drinks and over 4 more gallons of bottled water per year as compared with the households below 130 percent of poverty. Above poverty households also purchased more 2-percent, 1-percent, and skim milk, while households below 130 percent of poverty purchased more unflavored whole milk.

Household Size

The household size demographic has nine categories ranging from one household member to nine or more (fig. 34, data table 17) and includes average purchases by household size for those that bought. No household had more than nine members with the mean household size in the panel being 2.57 members. The largest category was the household size of two that had 2,704 observations of the 7,195 households in the data set.

As household size increases, purchases, on average, typically increase. This finding is largely due to the fact that the data deal primarily with food-at-home purchases. As family size increases, the household is less apt to dine

out or eat away from home for budgetary reasons. Every beverage listed is consumed in greater amounts in households with two or more persons compared with single-person households. The exception to this observation was frozen fruit drinks. Single-person households are either eating more on the go or away from home than multi-person households. As household size increases, powdered soft drinks, milk, and carbonated soft drinks are more heavily consumed at home.

Female Head of Household

Three demographics concerning the female head of household—age, employment status, and education level—were looked at next. It is assumed that the female head is largely responsible for food-at-home purchases. Six hundred seventy-one of the households had no female head of household or the household gave no information regarding age, employment, or education of a female head.

Age of Female Head of Household

There are eight categories of age for female head of households (fig. 35, data table 18). Households with the female head under 25 years old bought more powdered soft drinks than did all remaining households with female heads that are older. Households with older female heads bought considerably more coffee than did households with younger female heads. Coffee purchases ranged from 15.55 gallons for households with female heads under age 25, to 50.82 gallons for households with female heads older than 65. Carbonated soft drink purchases for households with female heads ages 40-44 was the highest level, 68.92 gallons. This figure is 17.05 gallons above the overall average of all households in the surveyed panel. Milk purchases also varied for any age of female-headed households starting at 42.19 gallons for those under 25 and then dropping to 31.19 gallons for those in the 25-29 bracket. From this level it slowly increased until the female head turned 45, then the average household purchase of milk decreased thereafter.

Employment of Female Head of Household

There are four categories of employment, ranging from not employed to three different categories of hours worked per week (fig. 36, data table 19). The majority of the beverage consumption changes little from one classification to the next. One notable difference is the purchase of tea for households where the female head worked 35-plus hours. The average purchase is 2 gallons less than households with different-aged female heads. Households with unemployed female heads bought more coffee for at-home consumption than did households with employed female heads. The average consumption is 49.46 gallons per year for unemployed female heads. The unemployed and fully employed female-head households bought less carbonated soft drinks on average than female heads who work part-time. Lastly, households that contain a female head who works fewer than 30 hours per week bought more milk on average than did other households, purchasing 40.7 gallons per year.

Education of Female Head of Household

There are six categories for education ranging from grade school education to post-college education (fig. 37, data table 20). There were 2,187 of the households in the data set, including a female head that attained some college education, followed by 1,821 households where the female head had graduated from college. Households with higher educated female heads bought more apple juice, orange juice, and other fruit juices than did households with less educated female heads. The average purchase of juices gradually increased as the education level of the female head rose. Conversely, powdered soft drink purchases per household decreased as education level increased, ranging from 20 gallons for households where the female head had some high school education to 15 gallons for households of female heads that attained a post-college education. Coffee and carbonated soft drink purchases decreased for households where the female heads in the households were more educated, similar to powdered soft drinks. This finding also was true for milk; purchases in households decreased as the education level of the female head in those households rose.

Race

The demographic for race had four categories: White, Black, Asian, and Other (fig. 38, data table 21). In the panel data, 83.5 percent of the households were White. Asian households bought more ready-to-drink fruit juices and orange juice than households of other race classifications did.

Consuming only 27.96 gallons of carbonated soft drinks per year, Asian households drank substantially fewer gallons when compared with White and Black households, who consumed 54.14 gallons and 35.51 gallons, respectively, on average per year. Black households bought more powdered soft drinks and ready-to-drink fruit drinks and less tea than did households of other races. White households purchased the greatest amount of coffee per year, 45.1 gallons compared with other races. White households also purchased the largest amounts of milk, but less bottled water on average than did households of different races.

Region

Four regions were studied: East, Central, South, and West (fig. 39, data table 22). Households located in the East bought more ready-to-drink fruit juices, orange juice, tea, and coffee than did households from other regions. Households located in the East and South purchased the least milk of any region, at about 32 gallons per year per household. Households located in the Central region bought more milk, carbonated soft drinks, and powdered soft drinks than did other household regions on average. Southern households purchased high levels of powdered soft drinks, though slightly less than did Central region households. Southern households also bought high levels of carbonated soft drinks. Households located in the West purchased more gallons of bottled water per year than did households located in other regions. Western households bought less orange juice and tea than did households from other regions.

Hispanic Origin (Ethnicity)

The data indicated that 457 of the 7,195 households in the panel were of Hispanic origin (fig. 40, data table 23). Hispanic households bought more ready-to-drink fruit drinks, powdered soft drinks, carbonated soft drinks, bottled water, and milk than did households that were not Hispanic. Hispanic households purchased less tea and coffee than non-Hispanic households. Hispanics purchased more milk overall than non-Hispanic households, with the majority of that milk being “whole” and 2-percent milk. Non-Hispanic households purchased more 1 percent and skim milk in contrast to Hispanic ones. Households of Hispanic origin bought more frozen concentrated orange juice than did non-Hispanic households.

Seasonality

The purchases of nonalcoholic beverages in the data set allowed seasonality to be analyzed (fig. 41, data table 24). Overall, the number of households that bought nonalcoholic beverages during all four quarters in 1999 and the average purchases of each nonalcoholic beverage were relatively stable. The purchases of carbonated soft drinks were slightly higher during the second and third quarter (the warmer months) than in other quarters. Milk purchases decreased slightly in the third and fourth quarters, relative to other quarters of the year. Coffee purchases were greatest in the fourth quarter (the holiday months) at 15.25 gallons, relative to other quarters of the year. Powdered soft drink purchases were the most seasonal beverage, with the number of households purchasing powdered soft drinks almost double for the second and third quarters, which includes the summer months when children are out of school.