February 2007

Homescan Data Description

Excerpted from:
ERS Data Quality Study Design
Contract 53-3K06-06-1301
Final Report

Prepared for

Ephraim Leibtag
U.S. Department of Agriculture
Economic Research Service
1800 M Street NW, Room N2124
Washington, DC 20036

Prepared by

Mary K. Muth
Peter H. Siegel
Chen Zhen

RTI International
Health, Social, and Economics Research
Research Triangle Park, NC 27709

RTI Project Number 0210153.001

Description of the Homescan Data

Scanner data collected on consumer purchases fall into two categories: point-of-sale or store scanner data and household-based scanner data. Point-of-sale scanner data are collected at cash registers and identify the products, quantities sold, and prices paid. Household-based scanner data come from a sample of households that scan universal product codes (UPCs) of all purchased products after each shopping trip. Although point-ofsale scanner data have been available to academic researchers since at least the early 1980s (e.g., Guadagni and Little [1983]), household-based scanner data are a more recent innovation. These data provide information on household demographic characteristics that are not available in store scanner data. In addition, because the household scanner data panelists are instructed to scan all purchases from all outlets, the datasets from household-based scanner data are more complete than datasets of purchases of individual households collected through loyalty card users. The latter data collection does not include information on household demographics and is likely subject to more measurement errors because of infrequent use of loyalty cards or use of someone else's card for convenience.

Nielsen and Information Resources Inc. (IRI) are the two major commercial suppliers of scanner data. The store scanner data service provided by Nielsen is called Scantrack; the one by IRI is called InfoScan. The in-home household scanner data collected by Nielsen is called Homescan; the IRI collection is called Consumer Network. Because the IRI in-home scanner data do not contain non-UPC random-weight perishable products, Homescan is the only option for obtaining these data. Nielsen started collecting in-home household scanner data in 1989. Both the number of the U.S. Homescan panelists and the number of projectable geographic areas have expanded substantially over the years. In 1989 the U.S. Homescan panel consisted of 15,000 households in two markets; in 2006 the panel had 125,000 households in 52 markets and 9 remaining areas in the 48 continental states. Table 2-1 provides a chronological account of major changes and additions to the Homescan panel and related services.

Although the entire panel scans all products with a UPC, a subset of the panel also records purchases of random-weight or non-UPC products (e.g., fresh fruits and vegetables, bakery products produced and packaged in the store, and meat products cut and packaged in the store). In 2004, the Fresh Foods panel comprised 15,000 households; this number has remained constant while the total panel has increased in the last couple of years.

In total, Homescan provides consumer information in 27 countries based on purchases from 265,000 households globally. For the U.S. Homescan panel, a household is provided a handheld scanner to record purchase information and upload all information on a weekly basis to Nielsen through a landline phone or Internet. In countries where using a handheld scanner is not feasible, households are instructed to collect product packages in a "dust bin." Nielsen field officers visit these households on a regular basis to scan the barcodes on these packages. In some countries, shopping diaries are also used.

Nielsen collects data from a larger number of households than are ultimately included in the static databases used for economic analysis. As explained in the next section, households must report data for at least 10 of 12 months during the year to be included in the static sample.

Table 2-1. Evolution of the Homescan Panel and Related Services

Year	Event	Number of Households	Number of Markets
1989	Initiated in-home scanner data collection with 15,000 households in two markets	15,000	2
1992	Added 25,000 households and 14 new markets	40,000	16
1994	Syndicated the Consumer Channel Facts supplying clients with statistics on channel shopper and category buyers within key retail channels and accounts	40,000	16
1995	Syndicated the Account Shopper Profiler, which identifies the demographics of shoppers by channel and account	40,000	16
1997	Added 12,000 households and three additional markets Started the Fresh Foods panel	52,000	19
1999	Added another 3,000 households and two additional markets Syndicated the Cross Outlet Facts, which provides information on shopping behavior of core and occasional shoppers across competitive channels and accounts Initiated the Hispanic panel in Los Angeles, CA	55,000	21
2000	Syndicated the Homescan Basket Facts and the New Product Facts	55,000	21
2001	Expanded the Los Angeles Hispanic panel Syndicated the Wal-Mart Consumer Insights	55,000	21
2002	Added 6,500 households and two new markets Started the Homescan Rx/OTC Panel intended to obtain information about consumer usage and purchase behavior for prescription and over-the-counter products	61,500	23
2004	Expanded MegaPanel by adding 29,500 households and five new markets	91,000	28
	Initiated the Homescan Online and the Panel Views Survey online		
	Introduced new scanner technology		
2005	Continued MegaPanel expansion (the number of households now at 125,000, number of markets now at 52 plus nine remaining areas)	125,000	52

Sources: Nielsen, Inc. May 2006. "Understanding the Homescan Advantage." Presentation by Liz Crews and Ed Groves, Nielsen, at RTI International, Research Triangle Park, NC. Harris, M.J. and N. Blisard. 1995. "Characteristics of the Nielsen Homescan Data." Working paper. Washington, DC: U.S. Department of Agriculture, Economic Research Service.

2.1

Each record of the Homescan dataset represents a purchase transaction by a household and includes the household demographic information and a projection factor for the household.

In the purchased data, households are included only if they participated in at least 10 of the 12 months during the year; these households are referred to as the "static" sample. For example, in 2004, approximately 40,000 households in the sample of 61,500 households provided data for the minimum number of months to be included in the static panel.

The 1998 through 2001 data are categorized into four datasets by food type:
- frozen foods and produce and meat products with a UPC (filename prefix: *fpm*),
- random-weight products without a UPC (filename prefix: *rw* or *random weight*),
- dairy products (filename prefix: *dairy* or *d*), and
- dry grocery products (filename prefix: *dg* or *drygrocery*).

The attached excel spreadsheet identifies the product groups encompassed in these datasets for 1998 through 2001. Note that only a subset of the regular Homescan panel (the Core panel), called the Fresh Foods panel, records purchases of random-weight foods without UPCs. Each record in the dataset contains the following variables for the products purchased:

- household ID number
- purchase date (YYMMDD)
- product module
- brand
- size
- multipack indicator (yes/no)
- UPC
- UPC description
- quantity
- price paid—deal
- price paid—nondeal
- coupon value
- flavor
- form
- formula
- container type
- salt content
- style
- type
- product
- variety
- store name identifier
- channel type identifier
- product group identifier
- department identifier

Table 2-2. ERS Homescan File Information, 1998–2001

File Name	File Contents	Number of Households	Number of Purchase Records
d1998.sas7bdat	Dairy, 1998	7,623	957,424
d1999.sas7bdat	Dairy, 1999	7,123	874,085
d2000.sas7bdat	Dairy, 2000	7,520	875,166
d2001.sas7bdat	Dairy, 2001	8,208	906,384
drygrocery1998.sas7bdat	Dry Grocery, 1998	7,624	4,508,518
drygrocery1999.sas7bdat	Dry Grocery, 1999	7,124	4,103,525
drygrocery2000.sas7bdat	Dry Grocery, 2000	7,523	4,073,924
drygrocery2001.sas7bdat	Dry Grocery, 2001	8,215	4,245,543
rw98.sas7bdat	Random Weight, 1998	7,623	1,390,082
rw99.sas7bdat	Random Weight, 1999	7,119	1,262,879
rw00.sas7bdat	Random Weight, 2000	7,496	1,331,080

rw01.sas7bdat	Random Weight, 2001	8,128	1,335,916
fpm1998.sas7bdat	Frozen Foods, Meat, and Produce, 1998	7,624	1,135,476
fpm1999.sas7bdat	Frozen Foods, Meat, and Produce, 1999	7,123	1,061,585
fpm2000.sas7bdat	Frozen Foods, Meat, and Produce, 2000	7,520	1,105,668
fpm2001.sas7bdat	Frozen Foods, Meat, and Produce, 2001	8,210	1,165,128

Appended to each record is the household demographic information as indicated in Table 2-3 and a projection factor for the household. Because the household ID is included in the dataset, researchers can track the same household over multiple years.

Table 2-3. Homescan Household Demographic Information

Field Title	Variable Description	Code Values
HHID	Household ID	7-digit ID number
HHSize	Household size	1 = single member; 2 = two members; 3 = three members; 4 = four members; 5 = five members; 6 = six members; 7 = seven members; 8 = eight members; 9 = nine+ members
HHInc	Household income	03 = under $5,000; 04 = $5,000–$7,999; 06 = $8,000–$9,999; 08 = $10,000–$11,999; 10 = $12,000–$14,999; 11 = $15,000–$19,999; 13 = $20,000–$24,999; 15 = $25,000–$29,999; 16 = $30,000–$34,999; 17 = $35,000–$39,999; 18 = $40,000–$44,999; 19 = $45,000–$49,999; 21 = $50,000–$59,999; 23 = $60,000–$69,999; 26 = $70,000–$99,999; 27 = $100,000 and over
AgeF	Age of female head	1 = under 25 years; 2 = 25–29 years; 3 = 30–34 years; 4 = 35–39 years; 5 = 40–44 years; 6 = 45–49 years; 7 = 50–54 years; 8 = 55–64 years; 9 = 65+ years; 0 = no female head
AgeM	Age of male head	Uses the same coding rules as AgeF
AC	Age and presence of children	1 = under 6 only; 2 = 6–12 only; 3 = 13–17 only; 4 = under 6 and 6–12; 5 = under 6 and 13–17; 6 = 6–12 and 13–17; 7 = under 6 and 6–12 and 13–17; 9 = no children under 18
MEmp	Male head employment	1 = under 30 hours; 2 = 30–34 hours; 3 = 35+ hours; 9 = not employed for pay; 0 = no male head
FEmp	Female head employment	Uses the same coding rules as MEmp

MEd	Male head education	1 = grade school; 2 = some high school; 3 = graduated high school; 4 = some college; 5 = graduated college; 6 = postcollege graduate; 0 = no male head or unknown
FEd	Female head education	Uses the same coding rules as MEd
Marital	Marital status	1 = married; 2 = widowed; 3 = divorced/separated; 4 = single; Blank = unknown
MOcc	Male head occupation	01 = professional; 02 = managers, officials; 03 = clerical; 04 = sales; 05 = craftsman/foreman (skilled); 06 = operative (semi-skilled); 07 = military; 08 = service workers and private household workers; 09 = farm owners, managers, foreman and laborers; 10 = students employed <30 hours; 11 = laborers; 12 = retired, unemployed
FOcc	Female head occupation	Uses the same coding rules as MOcc
HHComp	Household composition	1 = married; 2 = female head living with others related; 3 = male head living with others related; 5 = female living alone; 6 = female living with nonrelated; 7 = male living alone; 8 = male living with nonrelated;
Race	Race of the head of household[a]	1 = white; 2 = black; 3 = Oriental; 4 = other
Hisp	Whether the head of household is Hispanic[a]	1 = yes; 2 = no
Region	Census regions	1 = East; 2= Central; 3 = South; 4 = West;
MarketID	Scantrack markets	An integer ranging from 1 to 52 A missing value indicates the household is nonurban or suburban
Projection	Projection factor	Household-specific projection factor updated annually

[a] For households with two heads of household, Nielsen asks the respondent to designate which head of household makes most of the purchase decisions for the household and to indicate the race of that individual. If either of the two heads of household are Hispanic, the race of the household is designated as Hispanic.

The UPCs provide a description of the product including the brand name and short (abbreviated) brand description. However, the dataset does not include nutrition information, although in theory one could match the product with a product purchased in the store and record the nutrition information from the Nutrition Facts panel. The only types of product claim information included in the dataset are salt content and organic content. The organic variables—organic claim and USDA organic seal—allow researchers to distinguish organic foods from

nonorganic ones. These two variables are not included for random-weight purchases in any of the datasets. Thus, Huang and Lin (2006) used product descriptions in the random-weight datasets to identify organic tomatoes.

2.2 HOUSEHOLD SAMPLE RECRUITMENT

Nielsen defines the universe of households for the Homescan panel as all nongrouped (i.e., excluding institutions and dorms, for example) households residing in the 48 contiguous states. Information on household count (without contact information) by nine demographic variables at the county level was purchased from Claritas, Inc. (formerly Marketing Statistics). These demographic variables are

Households recruited for participation on the panel are first placed in a reserve pool and then selected for the panel based on the need to match geographic and demographic targets.

- household size (four levels),
- household income (four levels),
- household head age (four levels),
- female head education (four levels),
- male head education (four levels),
- the presence of children (yes/no),
- race (three levels),
- Hispanic (yes/no), and
- household head occupation (three levels).

The nine demographic variables result in 30 household count targets (each level in the bulleted list) that are used for the sample design and projection factor calculation. The Claritas data are supplied to Nielsen in October and used as the universe definition for January of the following year. The total number of U.S. households is forecasted on a monthly basis to avoid a step up between 2 years.

From the defined universe of households, Nielsen uses both direct mails and the Internet to recruit Homescan panelists. For the direct mail method, household names and addresses are purchased from suppliers such as Donnelly. The direct mail method is used in particular to solicit low-income or ethnic groups who may not have Internet access. For Internet recruitment, both banner ads at broad-reach Web sites and e-mail are used. The name of Nielsen is displayed in the Internet ads because it has high name recognition. About 30% of the solicitations are conducted using direct mail, and the other 70% are through the Internet. Because of the difficulty of recruiting and maintaining young, single-member, low-income and ethnic households, these types of households are recruited at an above-average rate. To further improve participation of the hard-to-recruit households, banner ads are placed on certain Web sites that these types of household ds are more likely to visit. However, these households are not offered different incentives to avoid potential distortions in purchasing behavior.

To ensure that the solicited households understand the implications (or workload) of panel participation, the panelist recruitment involves two phases:

- Respondents complete a two-sided single sheet questionnaire that contains primarily bubble-fill questions on demographic information.

. After returning the single-sheet questionnaire, a four-page booklet is sent out with questions on topics such as ownership of cars, pets, and level of income. Households that return clean questionnaires are put in the reserve pool.

Nielsen tries to provide a lot of information about the burden of being a panelist before selection because sample churn is expensive. Once in the reserve pool, the household waits at most 24 months for an opportunity to participate in the Homescan panel. To keep these households engaged while in the reserve, Nielsen sends out monthly surveys with short and "entertaining" questions. Typically, the response rate is 60% to 70% for online surveys and 75% to 80% for mail-in surveys. Note that the response rate is calculated based on the number of households that respond after the second reminder. It is possible that a household is never selected for panel participation during the 2-year period. In this case, the household is re-solicited for recruitment. About 17% of the households in this situation decide to rejoin the reserve pool and begin the waiting process again.

Nielsen selects replacement households from their reserve sample using a distance algorithm that identifies which households best match a target sample profile. The sample profile is based on 61 geographic areas (52 markets and 9 remaining areas) and the set of demographic variables.

Once on the panel, Homescan panelists are allowed to stay on the panel for as long as they want, unless violations in the recording of purchases are detected that would result in removal from the panel. The violator programs identify households that initially or over time cease to report reasonable levels of purchase activity. The length of panel participation varies greatly across households. As of August 2006, the average length of stay on the Core panel is a little less than 5 years; for the new Mega panelists, the average is about 8 months.

As of 2005, there were about 125,000 households in the Homescan MegaPanel, 61,500 of which belong to the Core panel that use the original scanner device. The smaller Fresh Foods panel consists of 15,000 households randomly selected from the Core panel. The Homescan MegaPanel covers 61 geographic areas (52 markets and 9 remaining areas), while the Fresh Foods panel participation is limited to 12 geographic areas.[1] A Homescan market corresponds to a Nielsen Scantrack market that is similar to a Metropolitan Statistical Area (MSA) and has a minimum of 1,500 panelists.

Being a Homescan panelist does not guarantee that a household's purchase information in a year will be included in the static database—the data that ERS purchased. As noted previously, to be included in the static sample, a household has to participate for at least 10 of 12 months. The difference between the total number of Homescan households and the number of households in the static sample can be substantial. For example, in the Fresh

[1] The Fresh Foods panel regions are Atlanta, Baltimore-Washington, Chicago, Los Angeles, New York, Philadelphia, San Antonio, San Francisco, remaining East Census region, remaining Central Census region, remaining South Census region, and remaining West Census region.

Foods panel data purchased by ERS for 1998 through 2003, approximately half of the 15,000 households were included in the static sample.

Because the household demographic variables are required for sample selection and calculation of projection factors, all queried demographic information is provided by the household. In addition, Nielsen asks panelists to update their demographic information annually and encourages certain demographic changes such as births to be updated quarterly.

2.3 **FOOD PURCHASE DATA COLLECTION PROCESS**

Households record price and quantity information for all products purchased each week. However, the process for recording or obtaining price and quantity information may vary depending on the store, the product, and whether discounts apply.

Once recruited on the panel, a household receives a handheld scanner and instructions in the mail. The household is asked to record all purchases made throughout each week. For all purchases, the household has to enter quantities purchased. Depending on where the purchase takes place, the household may or may not need to input price information into the scanner. For purchases at stores that are part of Nielsen's Scantrack, households do not need to enter prices of the products they purchase; instead, prices are imputed on a weekly basis by Nielsen using price information received from their Retail Marketing Service (RMS). The imputed price is a (volume) weighted average of discounted prices due to, for example, use of loyalty cards and regular prices. For purchases made at stores that are not part of Nielsen's Scantrack, households are instructed to enter the price using the handheld scanner. For purchases at either Scantrack or non-Scantrak stores, the household is also prompted to enter whether it was a discounted price or if the household used a coupon. Discounts are associated with each specific product.

Nielsen has routines to correct prices and quantities for multipacks based on knowing that a product or UPC is part of a multipack. It also has processes to adjust for other types of discounts. For example, a household bought two boxes of General Mills Cheerios at a promotional price of two for $5.00 from a non-Scantrack store. After getting home, the household would enter quantity of two and a total price of $5.00. The researcher could then obtain a unit price by dividing the total price by the number of units.
Because the in-home scanner does not have UPC information loaded, the product match-up is done after data are transmitted to Nielsen. In cases where Nielsen does not know what a UPC represents, Nielsen conducts postchecking. Nielsen field auditors are instructed to find the unmatched UPC in the store and document the barcode. However, not all unmatched products are eventually identified. According to Nielsen, all unmatched UPCs are placed in an unassigned (i.e., "999")

The data entry process differs for random-weight, fresh foods because the products do not have UPCs; the burden on households is substantially higher as a result.
category. Nielsen acknowledges that unmatched UPCs remain a challenge to the data collection. It is not possible to know the share of unmatched purchases in total reported

purchases since Nielsen does not include these unmatched products in the datasets sold to ERS; however, according to Nielsen, the percentage of records with unmatched UPCs is on the order of only 1% to 2%.[2] In any case, unmatched UPCs are one source of underreporting during the Homescan data collection.

For products consumed on the go, consumers are asked to bring the package home and scan it. During focus group meetings convened by Nielsen at different locations in the country, panelists claim that they scan these products. However, Nielsen acknowledges that they have no easy way to verify this claim.

For the Fresh Foods panel, the data-recording process is somewhat different. Nielsen does not have store-supplied price information for random-weight food products without UPCs. Therefore, it has to rely entirely on the Fresh Foods panelists for price information. This means a greater workload for the approximately 15,000 Fresh Foods panelists who have to scan UPC products and record quantity (or weight) and price information on non-UPC random-weight foods. To streamline the recording of non-UPC random-weight foods, a codebook with product descriptions is provided to each panelist for keying in the product codes. Depending on the product type, a household enters either quantity (or count) or weight. Nielsen has a process that converts reported quantities into weights. However, this conversion appears to be done on an as-requested basis.

2.4 PROJECTION FACTOR CALCULATION METHOD

When using sample data, the sum of weights in a subgroup is used to estimate the population count for the subgroup. That is, each sample household represents other households in the population, and a weight indicates how many households are represented. In the case of the Homescan data, the weights are the projection factors. Projection factors for the Homescan sample are developed using an iterative proportional fitting (IPF) procedure (also called a raking procedure) based on the Calmar software, which applies a linear algebra approach to creating household-level projection factors. IPF controls at the margins. IPF first forces the weighted sample totals of the levels of one variable to equal the population totals for that variable. Then, IPF forces the weighted sample totals of the levels of another variable to equal the population totals for that variable (Oh and Scheuren, 1983). This process continues for the total of nine demographic variables noted below:

. household size (four levels)
. household income (four levels)
. household head age (four levels)
. female head education (four levels)
. male head education (four levels)
. presence of children (yes/no)
. race (three levels)
. Hispanic (yes/no)
. household head occupation (three levels)

[2] We have examined the UPCs reported in these datasets. We find no "impossible" numbers such as 999. Further, there is no missing value for the product group variable, indicating every store purchase is assigned to a product group.

Household population (universe) count estimates are available at the county level for these demographic variables from Claritas, Inc. These estimates are updated on an annual basis at the beginning of each year, and the growth in the total U.S. household count is forecast on a monthly basis during the year.

Tables 2-4 through 2-8 show a simple hypothetical example to demonstrate how IPF works. Assume that there are only two demographic variables in the dataset: presence of children and Hispanic. Table 2-4 shows the number of sample households by these two variables.

Table 2-4. Distribution of Presence of Children by Hispanic

	Hispanic		
Presence of Children	Yes	No	Total
Yes	100	400	500
No	200	300	500
Total	300	700	1000

There is no sampling, base, or initial weight for the Homescan data, so an average projection factor is used initially to obtain weighted counts. Assuming a total population size of 1,000,000 gives an average adjustment factor of 1,000,000/1,000 = 1,000. Table 2-6 shows the weighted number of sample households by the two variables. The weighted number is computed as the sample size from Table 2-4 multiplied by the average projection factor of 1,000.

Table 2-5. Weighted Distribution of Presence of Children by Hispanic

	Hispanic		
Presence of Children	Yes	No	Total
Yes	100,000	400,000	500,000
No	200,000	300,000	500,000
Total	300,000	700,000	1,000,000

Let the known population totals for the two variables be as follows:
. Presence of children—Yes: 600,000
. Presence of children—No: 400,000
. Hispanic—Yes: 200,000
. Hispanic—No: 800,000

First, the row-weighted marginal totals are adjusted to match the row population marginal totals for presence of children. To do this, an index (ratio) is computed as the sum of the weights for the row divided by the known population count marginal total for the row. The original projection factors are divided by the index to obtain revised projection factors. For

example, the revised projection factor for the presence of children = Yes row is 1,000/(500,000/600,000) = 1,200. A revised weighted number is computed as the sample size from Table 2-5 multiplied by the revised projection factor. Table 2-6 shows the revised projection factors and the revised weighted counts.

Note from Table 2-6 that the weighted row totals now match the population totals for the presence of children. However, the weighted column totals do not match the population counts for Hispanic. Therefore, the procedure needs to be repeated by adjusting the column marginal totals to match the population counts for the marginal totals for Hispanic. The index is computed as the sum of the weights for the column divided by the known population count marginal total for the column. The revised projection factors from the row adjustment are divided by the index to obtain revised projection factors. For example, the revised projection factor for Hispanic = Yes column is 1,200/(280,000/200,000) = 857. A revised weighted number is computed as the sample size from Table 2-6 multiplied by the revised projection factor. Table 2-7 shows the revised projection factors and the revised weighted counts.

Table 2-6. Revised Projection Factors and Revised Weighted Distribution of Presence of Children by Hispanic After First Row Adjustment

	Hispanic				
	Yes		No		
Presence of Children	Revised Projection Factor	Revised Weight	Revised Projection Factor	Revised Weight	Total
Yes No Total	1,200 800	120,000 160,000 280,000	1,200 800	480,000 240,000 720,000	600,000 400,000 1,000,000

Table 2-7. Revised Projection Factors and Revised Weighted Distribution of Presence of Children by Hispanic After First Column Adjustment

	Hispanic				
	Yes		No		
Presence of Children	Revised Projection Factor	Revised Weight	Revised Projection Factor	Revised Weight	Total
Yes No Total	857 571	85,714 114,286 200,000	1,333 889	533,333 266,667 800,000	619,047 380,953 1,000,000

Note: Numbers in table are rounded.

After adjusting for both the row totals and column totals, an iteration has been completed. These calculations continue until the procedure converges (i.e., until the weighted row and

column totals both equal the population totals). In this example, it takes five iterations for convergence. Table 2-8 shows the final results.

Table 2-8. Final Projection Factors and Final Weighted Distribution of Presence of Children by Hispanic After Five Iterations

	Hispanic				
	Yes		No		
Presence of Children	Revised Projection Factor	Revised Weight	Revised Projection Factor	Revised Weight	Total
Yes	817	592	81,664	1,296	939	518,336	600,000
No			118,336			281,664	400,000
Total			200,000			800,000	1,000,000

Note: Numbers in table are rounded.

2.5 REFERENCES

Nielsen, Inc. May 2006. "Understanding the Homescan Advantage." Presentation by Liz Crews and Ed Groves, Nielsen at RTI International, Research Triangle Park, NC.

Guadagni, Peter M., and John D.C. Little. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science* 2:203-238.

Harris, M.J. and N. Blisard. 1995. "Characteristics of the Nielsen Homescan Data." Working paper. Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Huang, Chung L., and Biing-Hwan Lin. August 12-18, 2006. "A Hedonic Analysis on the Implicit Values of Fresh Tomatoes." International Association of Agricultural Economists Conference, Gold Coast, Australia.

Oh, H.L., and F.J. Scheuren. 1983. "Weighting Adjustment for Unit Nonresponse." In *Incomplete Data in Sample Surveys, Volume 2: Theory and Bibliographies*, W.G. Madow, I. Olkin, and D.B. Rubin (Eds.), pp. 143-184. New York: Academic Press.