Food Price and Diet and Health Outcomes

20 Years of the CARDIA Study

Kiyah J. Duffey, PhD; Penny Gordon-Larsen, PhD; James M. Shikany, MD; David Guilkey, PhD; David R. Jacobs Jr, PhD; Barry M. Popkin, PhD

Arch Intern Med. 2010;170(5):420-426

FOOD PRICES AND COST OF LIVING

Conducted quarterly across participating metropolitan and nonmetropolitan areas (approximately 300), the survey conducted by Council for Community and Economic Research (C2ER, formerly known as the American Chamber of Commerce Research Association) provides price variables for more than 60 consumer goods and services including grocery and restaurant food items, cost of living and overall price indexes, and cigarette prices. Although we would have ideally included the price of alternative types of milk (ie, low-fat milk) or away-from-home sandwiches (ie, chicken sandwich) in addition to the price variables used, this information was not collected and thus not available to us.

To account for inflation, we used the consumer price index (CPI), which represents changes in prices of all goods and services purchased for consumption by urban households, including user fees and sales and excise taxes, but excluding income taxes. The overall CPI is typically used in favor over food specific CPI values because it controls for total cost of living.

Price data were linked both temporally and spatially to Coronary Artery Risk Development in Young Adults (CARDIA) Study respondents, as described in the main article. While respondents lived in just 4 US metropolitan areas (representing 700 census tracts) at baseline, by year 20 they were located in 48 states, 1 federal district, 1 territory, 529 counties, and 3805 census tracts. Prices were imputed for individuals for whom there was not a direct match between residential location and city (defined as Metropolitan Statistical Area [MSA]) or year in which price data were collected. For example, if a respondent’s residential location had a single matching MSA code and price data were available for the year and quarter in which the respondent was surveyed, prices from that matching MSA were assigned to the respondent.

Information on cost of living (COL) was obtained from C2ER and spatially and temporally linked to each respondent via their residential location. The COL index is based on the following 6 components: housing, utilities, grocery items, transportation, health care, and miscellaneous goods and services. Data for the index were collected on more than 50,000 prices covering 60 different items by chambers of commerce, economic development organizations, or university-applied economic centers in each participating urban area.

STATISTICAL ANALYSIS

For analysis of price elasticity (the ratio of a percentage change in consumption to percentage change in price), we used 2-step marginal effect models. These models are useful for eliminating bias when examining outcomes where there are large proportions of zero values (ie, nonconsumers) that do not represent missing data and when the distribution of consumption is skewed. In the first step, the probability of consuming a particular food or beverage is estimated using a probit model with maximum likelihood estimation. In step 2, a log-log ordinary least square regression model is run conditioned on consuming the food or beverage in question (ie, uses only the subsample of consumers). Both models had the same specifications, adjusting for socioeconomic and demographic factors, logged values of other food and beverage items, logged COL, imputed price, and year. Examples of the soda models are as follows:

Step 1: \[ \text{Pr} (\text{AteSODA}) = \gamma_0 + \gamma_1 \ln (\text{Soda Price}) + \gamma_2 \text{Age} + \gamma_3 \text{Female} + \gamma_4 \text{Black} + \gamma_5 \text{Low Income} + \gamma_6 \text{Middle Income} + \gamma_7 \text{Education} + \gamma_8 \text{Single} + \gamma_9 \text{Married With Children} + \gamma_{10} \text{Single With Children} + \gamma_{11} \ln (\text{Food Prices}) + \gamma_{12} \ln (\text{COL}) + \gamma_{13} \text{Imputed Price} + \gamma_{14} \text{Year 0} + \gamma_{15} \text{Year 7}; \]

Step 2: \[ \ln (\text{SODAamt|AteSODA}) = \theta_0 + \theta_1 \ln (\text{Soda Price}) + \theta_2 \text{Age} + \theta_3 \text{Female} + \theta_4 \text{Black} + \theta_5 \text{Low Income} + \theta_6 \text{Middle Income} + \theta_7 \text{Education} + \theta_8 \text{Single} + \theta_9 \text{Married With Children} + \theta_{10} \text{Single With Children} + \theta_{11} \ln (\text{Food Prices}) + \theta_{12} \ln (\text{COL}) + \theta_{13} \text{Imputed Price} + \theta_{14} \text{Year 0} + \theta_{15} \text{Year 7}; \]

where \text{Pr} indicates probability; AteSODA, consumed from the SODA food group; COL, cost of living; Ln, natural log; and SODAamt, amount of soda consumed; and \( \gamma \) (gammmas) and \( \theta \) (thetas) represent the effect estimates for probit and ordinary least square regression models respectively.
To derive unconditional predicted logged consumption, the probability of consumption estimated in step 1 is multiplied by the expected logged energy intake estimated in step 2 in the following way:

$$E(\ln Y|X) = \Phi(X'\alpha) \times (X'\beta);$$

where $E$ indicates the expected marginal effect value for the natural log of outcome ($Y$) given the set of covariates ($X$) and $\Phi$ represents the standard normal cumulative distribution function; $X'$, the vector of explanatory variables; $\alpha$, the vector of marginal effects from the probit estimation (step 1); and $\beta$, the vector of marginal effects from the ordinary least squares estimation (step 2). The resulting estimates, interpreted as elasticities, are weighted means of the association between changes in price with changes in consumption for the full sample. Marginal estimates and elasticities were all generated using 1000 (bootstrapped) replications.

REFERENCES