

# **The Role of Health Information on Fruits and Vegetable Consumption**

**Ram N. Acharya**

acharyar@asu.edu

7001 E. Williams Field Rd.  
Mesa, AZ 85212

Selected Paper  
American Agricultural Economics Association Annual Meeting  
Chicago, IL, August 5 - August 8, 2001

Morrison School of Agribusiness and  
Resource Management  
Arizona State University East

May 15, 2001

---

Copyright 2001 by Ram N. Acharya. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

## **The Role of Health Information on Fruits and Vegetable Consumption**

A large body of scientific research have shown that the risk of cancer and other chronic diseases can be substantially reduced by embracing a diet high in fruits and vegetables and low in total and saturated fats. In particular, these studies show that diet plays an important role in determining the risk of chronic diseases such as coronary heart disease, cancer, stroke, diabetes, hypertension, and osteoporosis. The first four diseases alone are responsible in causing more than half of all deaths that occur in the United States. According to a recent USDA study, an improvement in dietary pattern alone could save us more than 70 million dollars each year in terms of medical costs, absenteeism, and premature deaths associated with these four diseases.

Realizing the gravity of the problem, both private as well as public institutions have been conducting various nutrition information and promotion programs. Some of the commonly known nutrition information or promotion programs include Five-a-Day for Better Health, Dietary Guidelines for Americans, Food Pyramid, and Surgeon General's Report on Diet and Health. Consumers are also receiving nutrition information from many other channels such as newspapers or magazine articles and advertisements, formal or informal extension courses, medical and health professionals, and nutrition labels. Despite these compelling scientific results and efforts made by various nutrition information and promotion programs, many Americans eat poorly.

In particular, an examination of the most recent health data from the Continuing Survey of Food Intakes by Individuals (CSFII) shows that the average daily servings of fruit differs considerably from the Food Guide Pyramid serving recommendations (McNamara et al.). Moreover, although the overall average estimates for vegetable groups are close to mean

recommendations of 3.5 servings, researchers caution that Americans may have to eat more servings of dark green and cruciferous vegetables to meet the recommended consumption levels (Johnston et al.). In this light, this study examines the role of health information on fruits and vegetable consumption using DHKS (Diet and Health Knowledge Survey) and CSFII survey data. A structural equation model is used to establish the link between theoretical constructs such as consumer's knowledge, awareness, and attitude towards the importance of eating healthy diet and their actual daily intake of fruits and vegetables. Results show that consumers who are knowledgeable, aware of diet related health problems, and those who think it is important to eat healthy diet tend to consume more fruits and vegetables.

### **Conceptual Framework**

Becker's household production theory and Grossman's health input demand function provide the basic theoretical framework for this study. Health is both produced as well as demanded by a consumer. A consumer demand for health input can be derived by maximizing his/her utility function subject to health production and resource constraint.

Health is a function of diet and other inputs such as consumer perceptions on diet. These perceptions are influenced by the information that a consumer has about the product. As a consumer receives new information about the health inputs (diet), he combines it with his experience to evaluate the product and forms new beliefs. Based on these beliefs, he develops new attitudes towards the product, which influence the demand for health inputs. In other words, as new health information is received, consumers revise their beliefs and attitudes towards diet and adjust their dietary decisions accordingly (Blaylock et al.).

The problem with this conceptualization of a consumer's dietary decision process is that the theoretical constructs such as beliefs and attitudes of an individual are not directly measurable. However, a structural equation model can be used to evaluate the relationships between such latent variables and a consumer's dietary decisions. The structural equation model exploits the relationships between these latent variables and their observed indicators to establish the link between the variables of interest and the theoretical constructs. Following Muthen, a general structural equation model can be specified as

$$\eta = \alpha + \beta\eta + \Gamma\xi + \zeta$$

and the measurement models are specified as

$$y = \Lambda_y\eta + \epsilon$$

$$x = \Lambda_x\xi + \delta$$

where  $\eta(mx1)$  is a parameter vector of intercepts;  $\beta(mxm)$  is the matrix of coefficients for the regressions among the endogenous variables ( $\eta_i$ ), which has zeros in diagonal and  $(I - \beta)$  is non-singular;  $\Gamma(mxn)$  is a matrix of coefficients of exogenous latent variables ( $\xi$ ) in the structural relationship; and  $\zeta$  is a random vector of residuals;  $y(px1)$  is a vector of observed response or outcome variables;  $x(qx1)$  is a vector of predictors, covariates, or input variables. The vectors  $\epsilon(px1)$  and  $\delta(qx1)$  are measurement errors in  $y$  and  $x$ , respectively. Since both of the latent variables ( $\eta$  and  $\xi$ ) are not observed, the observed response variables  $y$  and  $x$  are used to estimate factor loading ( $\Lambda_y$  and  $\Lambda_x$ ) on these latent variables (see Muthen; Joreskog and Sorbom for details).

## Empirical Model

Given the household income and market prices a representative household's demand for fruits and vegetables can be derived by maximizing the joint utility function subject to the household health production technology and family income. Assuming that the relevant functions satisfy regularity conditions, the demand functions for a household member can be expressed as (See Variam et al., for details)

$$\eta_i = \alpha_0 + \beta_1' \eta_i + \Gamma' \xi_i + \omega z_i + \zeta_i$$

and the measurement models can be expressed as

$$y_i = \alpha_{0y} + \lambda_{1y} \eta_i + \varepsilon_i$$

$$x_{ik} = \alpha_{0k} + \lambda_{1k}' \xi_i + \delta_{ik}.$$

where (assuming that USDA estimates for the number of daily fruit and vegetable servings are accurate)  $\eta_i = y_i =$  average daily servings of fruits ( $i=1$ ) and vegetables ( $i=2$ ). The latent exogenous variable ( $\xi_i = 1, 2, 3$ ) measures consumer's knowledge, awareness and attitude. All other variables and parameters, except for  $z_i$ , are defined as earlier.

Consumer food choices are influenced by various factors including consumer's beliefs, culture, environment, education, socio-demographic characteristics, economic status, as well as biological needs. For example, a recent study found that a less educated male living in a non-metropolitan area of Northeast region is less likely to consider dietary guidelines than his counterparts, while making his food choices (Nayga, 1999). A number of binary and continuous variables are included in  $z_i$  to account for socioeconomic and other factors. The binary variables included in the model are race (black, and minority other than black), sex (female), city, non-

MSA, employment status (employed), respondent's involvement in physical fitness activities (exercise), regions (northeast, mid-west, and west), participation in public nutrition programs (program), on diet due to health reasons (diet), vegetarian, smoker, education (holds post graduate degree), origin (Hispanic), and health status (self-reported health status - healthy). The continuous variables included in the model are annual family income, respondent's age, body mass index, and household size. The observed response variables used in the measurement model for  $x$  are reported in Table 2.

The data set used in this study comes from two sources. The information on the number of fruit and vegetable servings consumed by individuals come from the CSFII survey data and the information on these individual's level of nutrition awareness comes from the Diet and Health Knowledge Survey (DHKS) data set. In particular, the DHKS data set provides information about consumer's awareness on health problems associated with diet high on fat, cholesterol, sugar, salt and low on calcium and fiber, and the problem of overweight in general. The Dietary Guidelines for Americans recommends a daily consumption of at least two servings of fruits and three servings of vegetables per capita.

## **Results and discussion**

The goodness of fit indices are reported in Table 1. Although the model fails to meet goodness of fit test based on both the  $p$  value for the model chi-square test and the chi-square/degree of freedom ratio, it does well on other measures. However, the chi-square test is highly sensitive to sample size and is considered to be an excessively strict requirement, which is often violated in most applied situations (Hatcher). On the other hand, both the comparative fit index (CFI=0.98) and non-normed fit index (NNFI=0.96) are well above the recommended level of 0.9 indicating a

good fit of the model. Moreover, in all cases, the t-statistics for each factor loading is highly significant and the standardized factor loading are nontrivial in absolute magnitude (ranges between 0.34 to 0.72). The distribution of normalized residual is symmetrical with very few of them exceeding the absolute value of 2. These goodness of fit indices show that the model fits the data well.

We followed a two-step procedure as recommended by Anderson and Gerbing (1988). In the first step, confirmatory factor analysis was used to develop a measurement model that demonstrated an acceptable fit to the data. In the second step, the measurement model was modified and used to estimate the latent variable model. The measurement model describes the relationship between the theoretical constructs such as knowledge, awareness, and attitude with observed response variables. The DHKS survey questionnaire included a number of questions related to these constructs. In particular, there were ten questions related to consumer knowledge about nutritional content of a specific food item (see table 2). Similarly, there were seven and eleven questions relating to consumer awareness on diet related health problems, and consumer attitude toward healthy food choices, respectively.

Initially, all of these 28 variables were used to estimate three respective latent variables in the first stage. However, fifteen of them were dropped to maintain statistical consistency in the model. All other exogenous variables used in the model are observed. Therefore, unlike most other structural equation models, the model used in this study consists two different sets of exogenous variables - the latent variables which are not directly observed and observed variables such as the respondent's age and family income.

The estimated parameters of the structural model are reported in Table 3. One of the important factor determining an individual's demand for fruit and vegetable commodities is its own price. However, both of the data sets used in this study did not include price data. Since we are using cross sectional data and the regional prices of these commodities are expected to vary, variables representing these regions may reflect the impact of price variation across the region. Another important demand factor is an individual's family income. Considering that fruits and vegetables are normal goods, individual with higher income levels are expected to consume more fruits and vegetables. As expected, the coefficient associated with family income is positive and highly significant in both cases.

Other interesting results include a negative but significant coefficients for the variables females and smokers. Also people who are on diet seem to have higher level of fruit consumption and lower level of vegetable consumption. As expected people with higher body mass index have lower intake of fruits. Another interesting result seems to be the higher level of consumption for those who have post graduate degree, are employed, exercise more than twice a week, and consider themselves to be healthy.

A positive and significant beta coefficient in the vegetable equation shows that individuals with higher level of fruit consumption tend to consume more vegetables. As expected, the level of fruit consumption among minorities (other than Blacks and Hispanics) is significantly higher. Similarly, the fruit consumption level among the Hispanic population is significantly lower. As mentioned earlier, results show a significant variation in fruit consumption across the geographical regions. In particular, significantly higher level of fruit consumption is observed in the Northeast, the Midwest, and the West.



Finally, as expected, a positive relationship between the latent variables measuring consumer's knowledge, awareness, and attitude towards healthy eating behavior is observed. However, only two of these three latent variables hold significantly positive sign in the fruit equation. This result supports the hypothesis that health information as measured by consumer's knowledge on nutritional content on a particular food, awareness on diet related health problems, and attitude toward the importance of healthy eating behavior plays an important role in dietary decisions.

Table 1. Goodness of Fit Indices

Fit Function	0.1932
Goodness of Fit Index (GFI)	0.9891
GFI Adjusted for Degrees of Freedom (AGFI)	0.9822
Root Mean Square Residual (RMR)	0.1006
Parsimonious GFI (Mulaik, 1989)	0.6433
Chi-Square	815.6665
Chi-Square DF	387
Pr > Chi-Square	<.0001
Independence Model Chi-Square	18012
Independence Model Chi-Square DF	595
RMSEA Estimate	0.0162
RMSEA 90% Lower Confidence Limit	0.0146
RMSEA 90% Upper Confidence Limit	0.0177
ECVI Estimate	0.3093
ECVI 90% Lower Confidence Limit	0.2907
ECVI 90% Upper Confidence Limit	0.3297
Bentler's Comparative Fit Index	0.9754
Normal Theory Reweighted LS Chi-Square	815.9634
Akaike's Information Criterion	41.6665
Bozdogan's (1987) CAIC	-2802.1260
Schwarz's Bayesian Criterion	-2415.1260
McDonald's (1989) Centrality	0.9505
Bentler & Bonett's (1980) Non-normed Index	0.9622
Bentler & Bonett's (1980) NFI	0.9547
James, Mulaik, & Brett (1982) Parsimonious NFI	0.6210
Z-Test of Wilson & Hilferty (1931)	11.7978
Bollen (1986) Normed Index Rho1	0.9304
Bollen (1988) Non-normed Index Delta2	0.9757
Hoelter's (1983) Critical N	2247

**Table 2. Responses to health information questions and estimated parameter.**

	Variable	Parameter	t-Value
Knowledge ( $\xi_1$ )	Mean		
Which has more saturated fat (% of correct answers):			
Butter or margarine?	82.09	$\lambda_{K1}$ 1	
Skim milk or whole milk?	94.34	$\lambda_{K2}$ 0.528 **	15.70
Which has more fat:			
Peanuts or popcorn?	87.02	$\lambda_{K3}$ 0.667 **	14.60
Yogurt, or sour cream?	85.06	$\lambda_{K4}$ 0.776 **	15.34
Awareness ( $\xi_2$ )			
Have you heard about any health problem caused by:			
Eating too much fat?	88.01	$\lambda_{AW1}$ 1	
Eating too much salt or sodium?	89.37	$\lambda_{AW2}$ 0.857 **	22.29
Eating too much cholesterol?	90.84	$\lambda_{AW3}$ 0.892 **	23.55
Eating too much sugar?	80.11	$\lambda_{AW4}$ 0.738 **	16.66
Being overweight?	94.93	$\lambda_{AW5}$ 0.620 **	22.50
Attitude ( $\xi_3$ )			
How important is it to choose a diet			
Low in saturated fat?	3.41	$\lambda_{AT1}$ 1	
With plenty of fruits & vegetables?	3.61	$\lambda_{AT2}$ 0.838 **	38.67
Use sugars only in moderation?	3.37	$\lambda_{AT3}$ 0.822 **	33.06
With adequate fiber?	3.38	$\lambda_{AT4}$ 0.984 **	38.62

Note: Responses to the question relating to knowledge and awareness variables are converted to zero one scale where 1 represents correct answer and 0 otherwise. The attitude variables are measured in 4 point scale where 1=not at all important and 4=very important.

**Table 3. Structural Equation Parameters**

Exogenous Variables		Fruits	t-Value	Vegetables	t-Value
Fruits	$\beta_{21}$			0.104 **	5.41
<b>Knowledge</b>	$\gamma_1$	0.429 *	1.84	0.205	0.71
<b>Awareness</b>	$\gamma_2$	0.298	1.24	0.411	1.37
<b>Attitude</b>	$\gamma_3$	0.344 **	5.90	0.040	0.55
Family Income	$\gamma_4$	0.005 **	4.25	0.005 **	3.45
Female	$\gamma_5$	-0.199 **	-3.80	-0.914 **	-14.06
Vegetarian	$\gamma_6$	0.201	1.36	0.133	0.73
Smoker	$\gamma_7$	-0.549 **	-9.86	-0.109	-1.57
Age	$\gamma_8$	0.006 **	3.17	-0.003	-1.02
Employed	$\gamma_9$	0.283 **	-4.91	0.247 **	3.44
Exercise	$\gamma_{10}$	0.142 **	2.68	0.038	0.57
Post Graduate	$\gamma_{11}$	0.394 **	5.30	0.095	1.02
Healthy	$\gamma_{12}$	0.257 *	1.85	0.275	1.60
On Diet	$\gamma_{13}$	0.250 **	3.82	-0.186 *	-2.28
Body Mass Index	$\gamma_{14}$	-0.019 **	-3.99	0.003	0.52
Household Size	$\gamma_{15}$	-0.080 **	-4.49	-0.041	-1.84
Origin	$\gamma_{16}$	-0.193 *	1.92	-0.085	-0.68
Minority	$\gamma_{17}$	0.334 **	2.98	0.169	1.22
City	$\gamma_{18}$	0.069	1.20	-0.033	-0.46
Non-MSA	$\gamma_{19}$	-0.214 **	-3.50	0.055	0.72
North East	$\gamma_{20}$	0.272 **	3.95	-0.046	-0.54
Mid West	$\gamma_{21}$	0.164 **	2.58	0.136	1.72
West	$\gamma_{22}$	0.332 **	4.84	-0.024	-0.28
Program Participant	$\gamma_{23}$	0.330	0.76	0.208	0.39

Note: All observed exogenous variables except for family income, age, body mass index, and household size are binary variables. The dependent variables are the average number of fruit and vegetables servings (averages from the Day One and Day Two intake questionnaires). The total sample size is 4223. The overall model Bentler's Comparative Fit Index is 0.9754 and Bentler & Bonett's (1980) Non-normed Index is 0.9622.

\*\*, \* Denote significant at 1 and 5 percent level.

## References

- Anderson, J. C. and D. W. Gerbing. "Structural Equation Modeling in Practice: A review and Recommended Two-Step Approach." *Psychological Bulletin* 103(1988):411-423.
- Becker, G. S. "A Theory of the Allocation of Time." *Econ. J.* 75(September 1965):493-517.
- Blaylock, J., D. Smallwood, K. Kassel, J. Variam, and L. Aldrich. "Economics, food choices, and Nutrition." *Food Policy* 24(1999):269-86.
- Hatcher, L. *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*. SAS Institute Inc. North Carolina, 1998.
- McNamara, P. E., C. K. Panney, L. S. Kantor, and S. M. Krebs-Smith. "The Gap between Food Intakes and the Pyramid Recommendations: Measurement and Food System Ramifications." *Food Policy* 24(1999):117-33.
- Muthen, B. "Latent Variable Structural Equation Modeling with Categorical Data." *Journal of Econometrics* 23 (1983):43-65.
- Nayga, R. M., Jr. (1995). "Determinants of U.S. Household Expenditures on Fruit and Vegetables: A Note and Update." *Journal of Agricultural and Applied Economics* 27(December 1995):588-94.
- Nayga, R. M., Jr. (1997). "Obesity and Heart Disease Awareness: A Note on the Impact of Consumer Characteristics Using Qualitative Choice Analysis." *Applied Economics Letters* 4(4):229-31.
- Variyam, J. N., J. Blaylock, and D. Smallwood. "A Probit Latent Variable Model of Nutrition Information and Dietary Fiber Intake." *Amer. J. Agr. Econ.* 78(August 1996):628-39.