Health Knowledge and Consumer Use of Nutritional Labels: The Issue Revisited

Sung-Yong Kim, Rodolfo M. Nayga, Jr., and Oral Capps, Jr.

The role of health knowledge in consumer use of nutritional labels on food packages is explored using data from the 1995 Diet and Health Knowledge Survey. Two types of label use models, a binary choice label use model and a level of label use model, are employed with particular attention given to the endogeneity of health knowledge. The binary choice model is concerned with factors affecting the probability of label use. The level of label use model deals with factors affecting the number of food products in which label use occurred. The results show that health knowledge has a significant role in increasing label use.

Most diets of Americans still fall short of the *Dietary Guidelines for Americans* (Frazao). Several scientific reports also suggest that a large percentage of health risks is related to diet and lifestyle choices. For example, the National Cancer Institute estimates that 35% of cancer deaths are linked to diet. There are almost 1.4 million new cancer cases in the United States each year, and more than 500,000 deaths from this disease. With rising scientific knowledge about the role that dietary choices play in preventing diseases, simple dietary changes could have a significant impact on reducing disease rates.

In an effort to make information available and to teach consumers how to use nutritional information, the Nutrition Labeling and Education Act (NLEA) now requires nutritional labels on food packages. The objective is to provide consistent, understandable, and usable labels that can help consumers make healthier food choices. The implementation of the NLEA is estimated to cost food processors between $1.4 billion and $2.3 billion over the next 20 years (U.S. Food and Drug Administration). Even though the benefits to public health—measured in monetary terms—are expected to well exceed the cost, the benefits are conditional upon consumer use. It is possible that many individual diets fall short of the dietary guidelines because they do not use nutritional labels on food packages to help them in their buying decisions. Increasing consumer use of nutritional labels is important since it has been reported to improve the quality of consumers' diet (Kim et al.; Kim, Nayga, and Capps). Therefore, improving label use can have important public and health policy implications because of the benefits that improved diets can provide the society in general in terms of lives saved and reduction of health care costs. Nayga (1996) suggested that individuals armed with proper health knowledge are more likely to utilize nutritional labels in their food purchase decisions.

Several empirical studies have been conducted to evaluate the determinants of nutritional label use (Wang, Fletcher and Carley; Nayga 1996, 2000; Guthrie et al.; Klopp and McDonald; Nayga, Lipinski, and Savur). Some of these studies have recognized the importance of health knowledge on food purchase behavior. Nevertheless, little is known about the relationship between health knowledge and label use. Klopp and McDonald, and Wang et al. ignored the possible role of health knowledge in label use. Guthrie et al., and Nayga (1996) included nutrition information in their studies but it was treated as an exogenous variable. Such an approach precludes an understanding of the relationship between health knowledge and label use, thus opening up the results to potential simultaneity bias in estimating label use.

The authors are graduate assistant, associate professor, and professor, respectively, with the Department of Agricultural Economics at Texas A&M University. The authors thank anonymous journal referees for helpful comments and suggestions.

Agricultural and Resource Economics Review 30/1 (April 2001) 10–19

This paper revisits the role of health knowledge in nutritional label use. A direct measure of health knowledge (i.e., knowledge about linkage between diet and disease) is generated from responses to questions about the correlation between diets and diseases. Health knowledge is incorporated into the label use model as an explanatory variable. The health knowledge and label use variables are treated as endogenous variables and are jointly estimated. In addition, two types of label use model-a binary choice label use model and a level of label use model, are employed to fully capture behavior on the use of nutritional labels by consumers.

Model Specification

Consumers choose foods within the context of a total diet in order to obtain greater utility from their food. Part of that utility is derived from using food to maintain or improve health (Variyam, Blaylock, and Smallwood). However, consumers with different diet-health knowledge may choose different bundles of foods. Consumers may also have difficulty assessing the quality of foods they purchase even after consumption. In this sense, nutritional labels make it practicable for consumers to judge the nutritional quality of a food product before purchasing (Caswell and Mojduszka).

Nutritional label use and health knowledge are considered to be inputs used to produce healthier diets. The role of diet-health knowledge in label use is through the consumer's perceived marginal product of improved diet. Based on this perceived marginal product, the household decides whether or not to use nutritional labels by comparing the marginal benefit of improved quality of diet to the marginal cost of using a nutritional label.[1] Following Stigler's net benefit approach, the extent of a consumer's health knowledge will be determined by factors that affect the expected value or costs of acquiring knowledge. What factors are important is an empirical question that is addressed below.

The empirical models are comprised of a reduced form equation (1) of health knowledge and two structural equations of label use (2) and (3):

$$(1) \qquad I_i = f(\boldsymbol{X}_i, v_i).$$

$$(2) \qquad Pr(y_i^* > 0) = g(I_i, \boldsymbol{Z}_i, u_i).$$

$$(3) \qquad [Y_i \mid \Psi_i, y_i = 1] = h(I_i, \boldsymbol{Z}_i, \boldsymbol{G}_i, \varepsilon_i).$$

[1] The marginal cost of using label is not just the opportunity cost of time to read it, but also the foregone consumer surplus of lost consumption enjoyment.

Equation (1) describes a health knowledge equation where individual i's health knowledge I_i is represented by the number of correct knowledge about linkage between diet and disease. Variables in $\boldsymbol{X}$ include socio-economic characteristics, health status, source of health knowledge, and food stamp program participation (Kenkel 1990, 1991; Gould and Lin; Variyam, Blaylock, and Smallwood). Household income may indicate human capital beyond that given by formal education, and thus may reflect greater efficiency in processing information. However, higher income may also reflect higher opportunity cost of time, which may reduce time spent seeking out health knowledge. Education is used to proxy for sources of health knowledge, since well-educated people are both exposed to more information, and are better able to understand and process it (Kenkel). Employment may reflect the value of time and the cost of gathering health knowledge for the household (Becker; Ippolito and Mathios). Household size is likely to impact both intra-household allocation of resources and the time allocation, thus influencing the formulation of health knowledge (Gawn, Innes, and Rausser). The health status variable and age are included to reflect differences in the incentives to gather information. Racial, urbanization and regional differences may reflect differences in media exposure (Putler and Frazao). Either the Poisson or negative binomial distributions are typically used with the health knowledge model (Cameron and Trivedi). In this study, the count random variable I_i in equation (1) is assumed to follow the Poisson distribution with $v_i \sim N(0, \sigma_v^2)$.

Equation (2) is a binary choice label use model, which describes the probability of the ith consumer using the nutritional label. y_i^* indicates the difference in the utility with versus without label use. Thus, if the utility associated with label use is greater than the utility without label use (i.e. $y_i^* > 0$), then we observe that $y_i = 1$, that is, the ith consumer uses the nutritional label. Otherwise we observe that $y_i = 0$ (i.e., $y_i^* < 0$), that is, the ith consumer fails to use the nutritional label. u_i follows a standard normal distribution. The vector $\boldsymbol{Z}$ indicates exogenous variables and includes individual characteristics (age, sex and income), situation variables (employment, household type) and marketing environment (region of residence, urbanization) of the consumers (Guthrie et al.; Nayga 1996; Wang, Fletcher, and Carly). The education variable is excluded from $\boldsymbol{Z}$ because education's main role is assumed to be through health knowledge (Kenkel 1990). This specification is consistent with the work of Grossman and Michael.

Equation (3) represents the level of label use model expressed by the number (Y_i) of food products on which consumers use the nutrition label, conditional on any label use. Each individual's decision on label use is likely to differ among various food products because the perceived marginal benefit of label use varies by product. Even though consumers usually use nutritional labels, they might not be willing to use labels for certain food products. This observation suggests that the binary choice indicator of label use employed in previous studies may be limited in fully capturing behavior on the use of nutritional labels by consumers. Hence, the number of products where a particular individual uses the label is employed as another measure of label use, in addition to binary choice label use. The count random variable Y_i in equation (3) is assumed to follow the Poisson distribution with $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. $\boldsymbol{Z}$ and $\boldsymbol{G}$ vectors represent exogenous variables in the model. Ψ represents the union of I, $\boldsymbol{Z}$, and $\boldsymbol{G}$. The vector $\boldsymbol{G}$ includes variables that indicate how easily consumers access and process the information on the label (Schmidt and Spreng).

Estimation Procedure

The estimation technique for the standard count data model is used to estimate equation (1). For an endogenous regressor I, the maximum likelihood estimator of the equation (2) is generally inconsistent. Thus, we use the two-stage conditional maximum likelihood estimator (TSCML) proposed by Rivers and Vuong.[2] Following Rivers and Vuong, the residuals (v) of equation (1) are obtained from the Poisson regression of I on X in the first stage, and in the second stage, the simple probit model is estimated with I, $\boldsymbol{Z}$, and v as explanatory variables.

In estimating equation (3), standard count data model estimation is inappropriate because the dependent variable, the number of food groups, is observed only for those individuals who use label. It may be possible that unobservable influences on binary label use are positively related to unobservable influences on the level of label use. Terza showed that in this case, equation (3) becomes

$$(4)\quad [Y_i \mid \Psi_i, y_i = 1] = e^{\beta'\Psi+\kappa}\left[\frac{\Phi(\gamma' w_i + \theta)}{\Phi(\gamma' w_i)}\right] + \eta = e^{\beta_0'\Psi}\,\tau_i(\theta, \gamma) + \eta,$$

where $w_i = [I_i, Z_i]$, $\theta = \rho\sigma_\varepsilon$, $\kappa = \sigma_\varepsilon^2/2$, $\rho = \mathrm{corr}(u, \varepsilon)$, and conditional on w, η is normally distributed error term and independent of μ and ε. Since $\partial\Phi(\gamma' w_i + \theta)/\partial\theta = \phi(\gamma' w_i + \theta) > 0$ and $\theta = \rho\sigma_\varepsilon$, if ρ is positive, it increases the mean of Y_i. The correction term $\tau(\theta, \gamma)$ is similar in nature to the inverse Mill's ratio of Heckman. As in Heckman, ignoring the correction term leads to omitted variable bias in the estimation of β.

Following Terza, a two-step estimation procedure is used to estimate equation (4). The first step is simply to estimate the selection equation (2) using all observations by maximum likelihood. For the endogenous regressor, the TSCML is again used to estimate equation (2). Then using results of the probit model, the inverse Mill's ratio is computed for observations for which $y_i = 1$. The second step, using the selected sub-sample, is to estimate β_0 and θ by the Nonlinear Instrumental Two Stage (NLIV) estimation method. The predicted value obtained from the Poisson regression of the equation (1) is used as an instrument. Figure 1 shows the linkages between the models, variables, and estimation procedures discussed above.

The Murphy and Topel correction method is used to adjust the asymptotic covariance matrices of β_0 because the sample selection correction term is based on an estimate from another model. The asymptotic covariance matrix for the two step estimator based on Murphy and Topel is

$$(5)\quad V_{\beta_o,\theta} = (G_1'\,G_1)^{-1}[H + G_2 V_r G_2'](G_1' G_1)^{-1}_{\,,}$$

where G_1 is the matrix whose typical row is

$$g_{1i} = \frac{\partial E[Y_i \mid \Psi_i, y_i = 1]}{\partial\begin{pmatrix}\beta_o\\ \theta\end{pmatrix}} = E[Y_i \mid \Psi_i, y_i = 1]\begin{pmatrix}\Psi_i\\ c_i\end{pmatrix},$$

$$\text{where } c_i = \left(\frac{\phi(\gamma' w_i + \theta)}{\Phi(\gamma' w_i + \theta)}\right),$$

$$H = \sum_i e_i^2\, g_{1i}\, g_{1i}',$$

where e_i is the ith residual from the NLIV estimation of the equation (4), V_γ is the estimated covariance matrix of the TSCML estimates, γ, and G_2 is the sum of cross products of G_{1i} and

$$g_{2i} = \frac{\partial E[Y_i \mid \Psi_i, y_i = 1]}{\partial\gamma} = E[Y_i \mid \Psi_i, y_i = 1]\left(c_i - \frac{\phi(\gamma' w_i)}{\Phi(\gamma' w_i)}\right)w_i.$$

[2] TSCML estimator has several advantages over alternative estimators. First it is easier to compute than Amemiya's GLS estimator, and in some cases asymptotically more efficient. Second, Monte Carlo simulation conducted by Rivers and Vuong showed that the TSCML estimator performs favorably relative to alternatives. Third, the TSCML procedure allows us to construct several simple endogeneity tests in the binary choice label model.

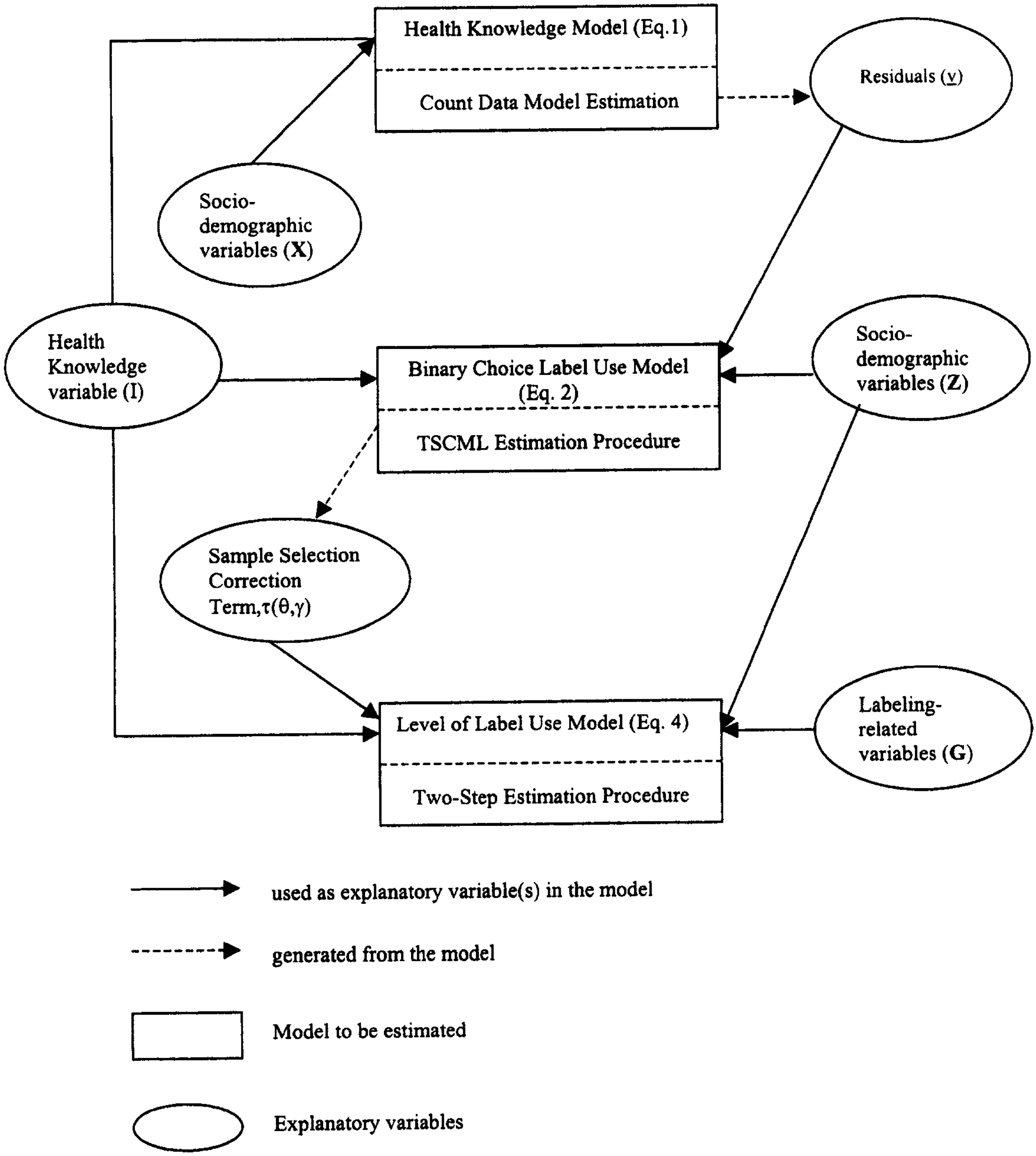

Figure 1. Framework for the Analysis

The coefficient β_{ij} in the level of label use model can be interpreted as the average proportionate change in $E[Y_i \mid \Psi_i]$ for a unit change in a particular variable j in Ψ for the ith consumer (Ψ_{ij}), unconditional on the consumer label use decision; that is, they measure the potential effect of a change in Ψ_{ij} on the sample, since

$$\frac{\partial E[Y_i \mid \Psi_i]}{\partial \Psi_{ij}} = \beta_{ij} E[Y_i \mid \Psi_i]. \tag{6}$$

Conditional effects on those who use labels also can be determined. Suppose there is a variable that appears both in w_i and Ψ_i, then

$$\frac{\partial E[Y_i \mid \Psi_i, y_i = 1]}{\partial \Psi_{ij}} = E[Y_i \mid \Psi_i, y_i = 1] \left[\beta_{ij} + \left(\frac{\phi(r'w_i + \theta)}{\Phi(r'w_i + \theta)} - \frac{\phi(r'w_i)}{\Phi(r'w_i)} \right) r_j \right]. \tag{7}$$

Equation (7) decomposes the effect of a change in Ψ_{ij} into two parts. The first part, β_{ij} is the direct effect on the mean of Y_i. The second part captures

Table 1. Definitions of the Variables

Variables	Description	Mean	Std. Dev.
Dependent Variables			
LBUSE	Respondent uses label when shopping (1 = yes; 0 = no)	0.808	
N_FOOD	Number of food product on which respondent uses label.	5.556	3.784
HEALTH KNOWLEDGE	Diet-disease knowledge (Index)	5.303	2.474
Explanatory Variables			
INCOME	Household Income (in 10,000 dollars)	3.455	2.548
AGE	Age of respondent (in years)	53.902	16.843
MALE	Respondent is male (1 = yes; 0 = no)	0.513	
B_RACE	Respondent is black (1 = yes; 0 = no)	0.110	
O_RACE	Respondent is other nonwhite races (1 = yes; 0 = no)	0.056	
EMPLOYED	Respondent is employed (1 = yes; 0 = no)	0.517	
CITY	Respondent resides in the central city (1 = yes; 0 = no)	0.273	
NONMETRO	Respondent resides in the non-metropolitan (1 = yes; 0 = no)	0.262	
HHSIZE	Number of the household members	2.522	
NORTHEAST	Respondent resides in the Northeast (1 = yes; 0 = no)	0.197	
WEST	Respondent resides in the West (1 = yes; 0 = no)	0.183	
MIDWEST	Respondent resides in the Midwest (1 = yes; 0 = no)	0.242	
FOOD STAMP	Respondent participate in the food stamp (1 = yes; 0 = no)	0.066	
SPECIAL DIET	Respondent has special diet (1 = yes; 0 = no)	0.182	
TV HOURS	Respondent watch TV for more than 5 hours (1 = yes; 0 = no)	0.185	
HEALTH STATUS	Respondent is in good health status (1 = yes; 0 = no)	0.496	
EDUCATION	Schooling in years	12.529	3.099
EASE	How much easy label is to understand (Index)	3.979	2.649
RELIABLE	How much reliable the description in the label is as the basis for choosing foods (Index)	3.069	2.490

Note: Base Group includes: white, female, unemployed, suburban, and south.

the indirect effect from label use decision that appears as a result of correlation between the unobserved components of Y_i and y_i.

Data

The data used are taken from the 1995 Diet and Health Knowledge Survey (DHKS) of USDA. The DHKS includes detailed information about the individual's socioeconomic background and questions on diet and health knowledge, addressing individual knowledge, awareness, and attitude on diet and health issues. The empirical work uses DHKS respondent files, providing sample size of 1760 observations.

The names, definitions, and means for principal variables are exhibited in table 1. Dependent variables include the decision on whether consumers use the nutritional labels when food shopping and the number of food products corresponding to consumer use of the nutritional label. About 80.8% of consumers in the sample use nutritional labels when food shopping.

The DHKS asked respondents about use of nutritional labels on 10 different types of food products: 'dessert items,' 'snack items,' 'frozen dinners or main dishes,' 'breakfast cereals,' 'cheese,' 'fruits or vegetables,' 'salad dressings,' 'table spreads like butter or margarine,' 'raw meat,' and 'processed meat products.' The second dependent variable Y reflects the number of food product items corresponding to an individual's use of a nutritional label when food shopping. For example, if consumers report using the label on dessert items like cookies, and on snack items, the value of Y is 2; if they report using the label on dessert items, snack items, and main dishes, then the value of Y is 3, . . . and so on. In the sample, the mean of this dependent variable is 5.6.

The explanatory variables used in the estimation of health knowledge and label use consist of consumer socio-demographic characteristics, food stamp participation, health status, TV hours, and special diet. Consumer characteristics include age in years, gender, household income, race, region, urbanization, education, employment status and household size.

The health knowledge variable used in this study is a measure of diet-disease knowledge. The diet-disease knowledge variable is generated from 17 selected questions asking respondents about the correlation between specific diseases and diet behaviors in the DHKS. An example question is: "Have you heard about any health problems caused by eating too much fat?" and "what health problems are these: cancer?" Those responding answers either "Yes" or "No." Each answer of "Yes" is a value of one, while each answer of "No" is given a value of zero. The variable I is the sum of the values of 0 and 1 corresponding the aforementioned 17 questions. On this basis, it is assumed that I reflects knowledge of the relationship between diet and disease. Since it reflects the answers to 17 questions, I can take values ranging from zero to 17 (complete knowledge). In this sample, the mean of I is 5.3. This knowledge variable can be interpreted as a measurement of an individual's perceptions of the parameters of the diet production function (Kenkel 1990).

Two variables are included in the level of label use equation to reflect consumer perceptions of the accessibility to and reliability of the information on labels (Schmidt and Spreng). One of the variables is "EASE," which reflects how easily consumers understand several types of nutrition information on labels and the other is, "RELIABLE," which reflects how reliable the consumer thinks the description in the label is as the basis for choosing foods. The variable, "EASE," is generated from 7 related questions in the DHKS such as "do you think the list of ingredient is very easy to understand, somewhat easy, or not too easy to understand?" The variable, "RELIABLE" is generated from 6 related questions such as "if a food label says a food is low-fat, would you say you are very confident, somewhat confident, or not too confident that the description is a reliable basis for choosing foods?"[3]

Empirical Results

The coefficient of the residuals (v) obtained from the OLS estimation of equation (1) provides a statistic that can be used to construct the test for the endogeneity in the binary choice label use model. The test of exogeneity in the level of label use model hinges on the procedure developed by Grogger. Following Grogger, the test statistic h is given by

$$h = \frac{(\hat{\delta}_{NLIV} - \hat{\delta}_{ML})^2}{[Var(\hat{\delta}_{NLIV}) - Var(\hat{\delta}_{ML})]} \tag{8}$$

where $\hat{\delta}$ and $var(\hat{\delta})$ are the coefficient estimate of the endogenous regressor I and its estimated variance in the level of label use model, respectively. Under the null hypothesis of the exogeneity in the health knowledge variable, the test statistics h asymptotically follows a chi-squared distribution with one degree of freedom.

The results of testing the exogeneity of health knowledge reveal a rejection of the assumption of exogeneity in the two label use models. The coefficient estimates of the residuals (v) of equation (1) is –0.2725 and significant at the 1% level (table 2). Thus the hypothesis that the health knowledge variable is exogenous in the binary label use model is rejected. Grogger's exogeneity test statistic is 5.3873. The corresponding chi-square test statistics is 3.84. Again the exogeneity of health knowledge is rejected in the level of label use model.

The empirical results of the health knowledge model are presented in column (1) while the label use models are exhibited in columns (2) and (3) in table 2. As expected, education and income are positively related to health knowledge. The estimated effect of income implies that health knowledge is a normal good. This result is consistent with those of Kenkel (1990), Gould and Lin, and Variyam, Blaylock, and Smallwood. Age, however is negatively related to health knowledge. Conflicting results were reported in other studies regarding the relationship between age and health knowledge. Kenkel (1990) reported a positive correlation between age and health knowledge, while Gould and Lin showed that age has no significant role in explaining differences in health knowledge. A possible explanation is that the incentive to gather knowledge increases with age. Alternatively, younger people have grown up in an era where health knowledge is more easily available than did older people implying lower information search cost (Gould and Lin).

The results from the health knowledge regression also indicate that females tend to be better informed than males. This result implies that females in general invest more on this type of human capital perhaps because women play an important role in the household production of family health (Sindelar). The levels of health knowledge are not significantly different among racial groups. Individuals on special diet are more knowledgeable about diet-disease relationships than those who are not on special diet. Also, individuals who perceive their health status to be better are more knowledgeable about diet-disease relationships than others. Results also indicate that individuals in the West tend to be more informed than individuals in the South, while individuals in the Northeast or Midwest are not likely more informed than individuals

[3] For more detail about the construction of the variables EASE and RELIABLE, see appendix.

Table 2. Estimation Results of Health Knowledge, and Label Use Models

Variables	(1) Health Knowledge Model	(2) Binary Choice Label Use Model	(3) Level of Label Use Model
Constant	1.1281**	−0.3955	1.3469**
	(12.989)	(−0.918)	(14.546)
INCOME	0.0306**	−0.0208	0.0010
	(7.342)	(−0.796)	(0.183)
AGE	−0.0021*	−0.0062*	−0.0018
	(−2.561)	(−2.002)	(−1.799)
MALE	−0.1320**	−0.5171**	−0.1449**
	(−6.220)	(−5.710)	(−4.834)
B_RACE	−0.0465	0.1327	0.0639
	(−1.221)	(1.043)	(1.481)
O_RACE	−0.0948	0.1019	0.0402
	(−1.904)	(0.558)	(0.717)
EMPLOYED	−0.0091	0.0464	0.0263
	(−0.354)	(0.484)	(0.921)
CITY	−0.0211	0.0027	−0.0475
	(−0.804)	(0.028)	(−1.668)
NONMETRO	−0.0371	−0.0455	−0.0618*
	(−1.401)	(−0.486)	(−2.060)
HHSIZE	−0.0072	−0.0432	−0.0082
	(−0.864)	(−1.489)	(−0.813)
NORTHEAST	0.0187	0.2230*	−0.0746*
	(0.630)	(2.019)	(−2.143)
WEST	0.0747*	−0.0244	−0.1053**
	(2.461)	(−0.202)	(−3.133)
MIDWEST	0.0423	0.0514	−0.0170
	(1.546)	(0.507)	(−0.586)
FOOD STAMP	0.0841	−0.2932	−0.1008
	(1.761)	(−1.915)	(−1.642)
SPECIAL DIET	0.1856**	−0.0774	0.0946**
	(7.057)	(−0.632)	(3.625)
TV HOURS	−0.0051		
	(−0.181)		
HEALTH STATUS	0.0823**		
	(3.681)		
EDUCATION	0.0422**		
	(10.005)		
HEALTH KNOWLEDGE		0.4123**	0.0616**
		(6.263)	(9.012)
EASE			0.0423**
			(6.136)
RELIABLE			0.0273**
			(5.069)
$\underline{V}$		−0.2725**	
		(−4.085)	
θ			0.5899**
			(3.239)
R^2	0.172	0.169	0.098
N	1760	1760	1422

Note: t-ratios are in parentheses; single asterisk indicates *t*-value is significant at .05 level; double asterisk indicates *t*-value is significant at .01 level; for the health knowledge equation, the R^2 reported is pseudo R^2 based on the Pearson residuals; for the binary label use equation, the R^2 reported is McFadden's pseudo-R^2 measure.

in the South. In addition, the results show that there is no significant difference in the level of health knowledge between food stamp participants and non-food stamp participants.

The regression results for binary label use equation (2) in column (2) of table 2 indicate that consumers with higher health knowledge are more likely to use nutritional labels when food shopping. This result is consistent with the argument that poorly informed consumers tend to underestimate the marginal product of label use.[4] The results for

[4] However, if a consumer does not know about the health implications of dietary choices, nutrition labeling could have a lower marginal value for him/her. Thus, the value of a particular information may also depend on other background information necessary to understand it.

Table 3. Endogeneity Effects of Health Knowledge in Label Use Models

	Probability of Using Label		The Number of Food Products	
Level of Health Knowledge	Endogenous Knowledge	Exogenous Knowledge	Endogenous Knowledge	Exogenous Knowledge
Low level				
0	0.1311	0.5905	4.4	5.5
1	0.2391	0.6492	4.7	5.7
2	0.3833	0.7046	5.0	5.8
3	0.5460	0.7555	5.3	6.0
4	0.7012	0.8013	5.6	6.1
5	0.8264	0.8415	6.0	6.3
High level				
6	0.9119	0.8759	6.4	6.4
7	0.9612	0.9048	6.8	6.6
8	0.9853	0.9284	7.2	6.7
9	0.9952	0.9472	7.7	6.9
10	0.9987	0.9618	8.1	7.1
11	0.9997	0.9730	8.7	7.2
12	0.9999	0.9813	9.2	7.4
13	1.0000	0.9873	9.8	7.6

the other explanatory variables in the binary label use model are generally consistent with prior expectations. The probability of label use decreases as age increase. Females are more likely than males to use food labels. These results are consistent with those of Guthrie et al.

The probability of using the label also varies depending on the region where an individual resides. Compared with individuals from the South, individuals in the Northeast region are more likely to use labels. However, Midwest or West region residents showed no difference from South region residents in terms of the probability of using nutritional labels. This finding is consistent with that of Nayga (1996). In addition, the results show that there is no significant difference in using nutritional labels between food stamp participants and non-food stamp participants.

Examining the regression results in the level of label use model (column 3 in table 2), we see that health knowledge has a significant and positive role in determining the level of label use model, based on the number of food products, conditional on label use. Thus, more informed individuals not only have higher probabilities of using labels when food shopping, but also have higher probabilities of using labels of more food product types. Older people use nutritional labels for a smaller number of food products than younger people. Females use nutritional labels for more food products than males.

Certain differences between the results of the two label use models are evident. For instance, urbanization is significant in the level of label use model but not in the binary choice label use model. Some differences in the regional results also are evident. Individuals in the South use nutritional labels of more food products than others in the Northeast or West region. In addition, individuals who are on a special diet use labels for more food products than others. Consistent with prior expectations, the "EASE" and "RELIABLE" variables are significant factors in the level of label use model. The correlation coefficient, θ, turned out to be positive and significant which means that failing to correct correlation between two types of label use models can lead to a bias in the mean of the level of label use, and hence a bias in the estimated health knowledge effect.

The treatment of health knowledge as an endogenous explanatory variable has important implications for the parameter estimates. Neglecting the endogeneity of I apparently results in downwardly biased estimates of I in equations (2) and (3). When equations (2) and (3) are estimated using the actual value for I rather than the predicted value, the coefficients are 0.1544 and 0.0231, respectively. They are still positive and significant, but smaller than that reported. Finally, to further understand the endogeneity bias effect in label use, we simulated the probability of using label and the number of food products for varying values of the health knowledge variable (table 3). The simulation results show the bias on the health knowledge effects for the two measures of label use. When endogeneity effects are not accounted for, the probabilities of label use are upwardly biased at below average levels of health knowledge and downwardly biased at above average health knowledge levels. Thus, the bias caused by neglecting endogenous health knowledge in the models perhaps can lead to erroneous policy or consumer education recommendations.

Summary and Conclusions

The role of health knowledge in nutritional label use is explored in this paper. Health knowledge and label use were assumed to be health inputs introduced into the household production function of family diet, and jointly estimated. This paper used a measure of health knowledge, instead of proxies, and estimated two types of label use models consisting of a binary label use model and the level of label use model. Particular attention was given to the endogeneity of health knowledge in the label use model ignored in previous studies.

The results revealed that health knowledge, as represented by diet-disease knowledge, has a significant role in increasing label use. This finding is important in this exploratory work because of its implications for consumer nutrition education programs. This finding is also more critical in light of Kim et al.'s and Kim, Nayga, and Capps' findings concerning the positive impact of label use on the quality of consumers' food intakes. Hence, public education and extension efforts should be targeted to those with lower health knowledge (e.g., lower income, males, less educated) to increase label use.

In addition, the results indicated that those who use nutritional labels tend to perceive the labels to be more reliable and easier to use. Making labels easier to use is a possibility but the reliability issue is a harder issue to address. However, it is possible that public education about the benefits of label use can increase consumers' trust and confidence on these labels (Nayga 1999).

References

Becker, G.S. 1965. "A Theory of the Allocation of Time." *Economic Journal* 75:493–517.

Cameron, A. and P. Trivedi. 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics* 1:29–53.

Caswell, J.A., and E.M. Mojduszka. 1996. "Using Informational Labeling to Influence the Market for Quality in Food Products." *American Journal of Agricultural Economics* 78:1248–53.

Frazao, E. 1995. *The American Diet: Health and Economic Consequences.* USDA Information Bull. No. 711. U.S. Department of Agriculture. Washington DC.

Gawn, G., R. Innes, and G. Rausser. 1993. "Nutrient Demand and the Allocation of Time: Evidence from Guam." *Applied Economics* 25:811–30.

Gould, B.W., and H.C. Lin. 1994. "Nutrition Information and Household Dietary Fat Intake." *Journal of Agricultural and Resource Economics* 19:349–65.

Grogger, J. 1990. "A Simple Test for Exogeneity in Probit, Logit and Poisson Regression Models." *Economics Letters* 33:329–32.

Grossman, M. 1972. "On the Concept of Health Capital and Demand for Health," *Journal of Political Economy* 80: 223–55.

Guthrie, J., J. Fox, L. Cleveland, and S. Welsh. 1995. "Who Uses Nutrition Labeling and What Effects Does Label Use Have on Diet Quality?" *Journal of Nutrition Education* 27:153–72.

Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–62.

Ippolito, D.M., and A.D. Mathios. 1990. "Information, Advertising, and Health Choices: A Study of the Cereal Market." *RAND Journal of Economics* 21:459–80.

Kenkel, D.S. 1990. "Consumer, Health Information and the Demand for Medical Care." *Review of Economics and Statistics* 72:587–95.

———. 1991. "Health Behavior, Health Knowledge, and Schooling." *Journal of Political Economy* 83:287–305.

Kim, S., R.M. Nayga, O. Capps, and B. Tepper. 1999. "Consumer Label Use and Diet Quality: An Endogenous Switching Regression Analysis." Invited paper presented at the Food and Agricultural Marketing Consortium Conference on *New Economic Approaches to Consumer Welfare and Nutrition,* Alexandria, Virginia.

Kim, S., R.M. Nayga, Jr., and O. Capps, Jr. 2000. "The Effect of Food Label Use on Nutrient Intakes: An Endogenous Switching Regression Analysis." *Journal of Agricultural and Resource Economics* 25(1):215–231.

Klopp, P., and M. McDonald. 1981. "Nutrition Labels: An Exploratory Study of Consumer Reasons for Nonuse." *Journal of Consumer Affairs* 15:301–16.

Michael, R. 1972. *Effect of Education on Efficiency in Consumption.* Occasion Paper 116. National Bureau of Economic Research.

Murphy, K.M., and R.H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business & Economic Statistics* 3:370–79.

Nayga Jr., R.M. 1996. "Determinants of Consumers' Use of Nutritional Information on Food Packages." *Journal of Agricultural and Applied Economics* 28:303–12.

———. 1999. "On Consumers' Perception about the Reliability of Nutrient Content Claims on Food Labels." *Journal of International Food and Agribusiness Marketing* 11:43–55.

Nayga Jr., R.M., D. Lipinski, and N. Savur. 1998. "Consumers' Use of Nutritional Labels While Food Shopping and At Home." *The Journal of Consumer Affairs* 32(1):106–120.

Nayga, Jr., R.M. 2000. "Nutrition Knowledge, Gender, and Food Label Use." *The Journal of Consumer Affairs* 34: 97–112.

Pulter, D.S. and E. Frazao. 1994. "Consumer Awareness of Diet-Disease Relationship and Dietary Behavior: the Case of Dietary Fat." *Journal of Agricultural Economic Research* 45:3–17.

Rivers, D. and Q.H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39:347–66.

Schmidt, J.B., and R.A. Spreng. 1996. "A Proposed Model of External Consumer Information Search." *Journal of the Academy of Marketing Science* 24:246–56.

Sindelar, J.L. 1982. "Differential Use of Medical Care by Sex." *Journal of Political Economy* 90:1003–19.

Stigler, G.J. 1961. "The Economics of Information." *Journal of Political Economy* 69:213–25.

Terza, J.V. 1998. "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects." *Journal of Econometrics* 84:129–54.

Variyam, J.N., J. Blaylock, and D. Smallwood. 1996. "A Probit Latent Variable Model of Nutrient Information and Dietary Fiber Intake." *American Journal of Agricultural Economics* 78:628–39.

U.S. Food and Drug Administration. 1995. *The New Food Label.* Publication No. BG. 95–12.

Wang, G., S.M. Fletcher, and D.H. Carley. 1995. "Consumer Utilization of Food Labeling as a Source of Nutrition Information." *Journal of Consumer Affairs* 29:368–80.

Appendix

1. Creating health knowledge variable *I*: A sample question is as follows: "Have you heard about any health problems caused by eating too much fat?" and "what health problems are these: cancer?" The following 17 questions in the DHKS are selected to construct the health knowledge variable *I*. These 17 questions describe all types of health problems caused by eating too much (or not eating enough) of specific nutrients.

(1) 5 types of health problems caused by eating too much fat:
KQ6_A_01: arteriosclerosis, coronary disease, heart attack, . . . , etc.
KQ6_A_05: cancer
KQ6_A_06: colon problems . . .
KQ6_A_12: high blood pressure . . .
KQ6_A_15: overweight, obesity . . .

(2) 2 types of health problems caused by not eating enough fiber:
KQ6_B_05: cancer
KQ6_B_06: colon problems . . .

(3) 2 types of health problems caused by eating too much sodium:
KQ6_C_01: arteriosclerosis, overweight, obesity . . .
KQ6_C_12: high blood pressure . . .

(4) 1 type of health problem caused by not eating enough calcium:
KQ6_d_03: bone problem

(5) 3 types of health problems caused by eating too much cholesterol:
KQ6_E_01: arteriosclerosis, coronary disease, heart attack, . . . , etc.
KQ6_E_12: high blood pressure . . .
KQ6_E_15: overweight, obesity . . .

(6) 2 types of health problems caused by eating too much sugar:
KQ6_F_12: high blood pressure . . .
KQ6_F_15: overweight, obesity . . .

(7) 2 types of health problems caused by being overweight
KQ6_G_01: arteriosclerosis, coronary disease, heart attack, . . . , etc.
KQ6_G_12: high blood pressure . . .

If each answer of above questions is "1" (i.e., "Yes"), then a value of one is given, while a value of zero is given to each answer of "0" (i.e., "No"). The variable *I* is simply the sum of scores on 17 questions. Thus, it has maximum value of 17, and minimum value of zero.

2. Creating the variable "EASE": There are 7 questions as to whether each of 7 types of information on the label is easily understood or not. An example question is as follows: "Do you think the list of ingredient is very easy to understand, somewhat easy, or not too easy to understand?" The 7 types of information illustrated in the DHKS question are (1) the list of ingredients, (2) the short phrase like "low-fat" or "light" or good source of fiber (3) the number of calories in a serving, (4) the number of calories from fat in a serving, (5) the number of grams or milligrams of nutrients, (6) the percent of the daily value for each nutrient, and (7) description like "lean" or "extra lean" on meats. If the response is "very easy" or "somewhat easy," then a value of one was given, otherwise, a value of zero. The variable EASE is simply the sum of score on these 7 questions. Thus, the variable EASE has a maximum value of 7 and a minimum value of zero.

3. Creating the variable "RELIABLE": There are 6 questions as to whether the description on the label is a reliable basis for choosing food. An example question is as follows: "If a food label says a food is low-fat, would you say you are very confident, somewhat confident, or not too confident that the description is a reliable basis for choosing foods?" The 6 descriptions on the label illustrated in the DHKS question are: (1) low fat food, (2) low cholesterol food, (3) a good source of fiber, (4) light food, (5) healthy food, and (6) extra lean food. If the response is "very confident" or "somewhat confident," then a value of one was given, otherwise, a value of zero. Thus, the variable RELIABLE has a maximum value of 6 and a minimum value of zero.