Changes in Factors Contributing to Rising Body Mass Index: 1997 versus 2002

Obesity has become an epidemic over the last decade. It is associated with increased health risk for chronic diseases such as heart disease, type 2 diabetes, high blood pressure, stroke, and some forms of cancer. Currently over 30% of American adults are obese, more than twice the percentage prevalent in 1980 (American Obesity Association). In this paper we explore the factors that contribute to rising body mass index (BMI) and obesity, and study the differences in years 1997 and 2002. We use a multilevel econometric approach to model both BMI and obesity as functions of individual behavior and external environment. We specifically discuss the results on BMI since obesity is derived from BMI itself. BMI is a continuous variable which is regressed against individual characteristics clustered within state factors. Significant differences are found between the two years mainly at the individual level. Almost all individual level predictors have significant association with BMI in year 2002, not so in 1997. Among several race categories, Black non-Hispanics and multiracial non-Hispanics have higher BMI than White non-Hispanics. Those who are unemployed have higher BMI than those employed. One puzzling result is that in 1997 smokers have higher BMI than non-smokers, whereas in 2002 this result is completely reversed. At the state level, for both 1997 and 2002, high unemployment rate and lower expenditure on food-at-home are associated with higher BMI. Two additional variables are significant in 2002 – expenditure on other-food at home (those rich in sugars and fats) and percentage living in metropolitan statistical areas. The former has positive association with BMI, whereas the later has a negative association.

Bidisha Mandal[1], The Ohio State University
Wen Chern[2], The Ohio State University

Introduction

Since the 1980s, Americans have been talking about muscle tone, exercise routines, and being in shape (Cassell). Innumerable fitness centers promote the importance of taking control of one's health. People who are overweight are considered unfit. Media reflects this view as overweight people are usually seen in character or supporting roles, and fashion models splashed across a plethora of magazines are always thin. However, obesity is more than just a cosmetic problem; it is the second leading cause of preventable death in the U.S., behind tobacco usage (McGinnis and Foege). Obesity is a complex chronic disease involving environmental, genetic, metabolic, and behavioral components. Genes, understandably, are hard to control, but how easy is it to eat right and exercise regularly? In an economic model, our weight is a rational choice along several dimensions – occupation, leisure time activity, residence, and of course, food intake (Philipson). In an agricultural society, physical activity is part of the occupation. But in a post-industrial developed society, like U.S., where most work is sedentary, one has to pay to stay in shape. Thus, the decline in work related physical activity seems to be one of the prime causes of obesity.

Another explanation may be provided by our increasing dependence on market-produced food as a substitute for household-produced food. Fast-food is often blamed for the rise in obesity. But, as the value of time increases, it is only natural to turn to food that is delivered faster. No wonder there is a higher demand for burgers than for the healthier sushi. Not only is this trend seen in the market, but also at home. Relatively inexpensive pre-cooked meals have flooded the grocery stores. So why spend half an hour over one meal, when it may be prepared in five minutes?

Kuchler and Golan investigated whether failure in food markets may help explain the growth of overweight and obesity in the United States. Given the constant onslaught of media coverage devoted to diet and weight these days along with information from physicians, government education programs, nutrition labels, and product health claims, it is difficult to believe that Americans are not conscious of the relationship between a healthful diet and obesity. Nevertheless, the authors did find existence of two important information blackout zones – public perceptions of appropriate weight, and information on food sold at restaurants and fast-food establishments. They found that among individuals whom professionals would classify as obese, 13% said that their weight is about right or even too low. Although the 1994 National Labeling and Education Act require that manufacturers disclose nutritional information on the label of almost all packaged food, it does not require the same for food purchased at restaurants. This information gap is vital since the nutritional content of food from restaurants tends to be less healthy than food prepared at home (Guthrie, Lin and Frazão). A 2003 Gallup Poll survey found that two thirds of consumers suspect that most food sold at fast-food restaurants was not good for them (Saad). However, consumers may not be able to gauge precisely the nutritional content of the food available in restaurants.

Science journalist Gary Taubes reports that the obesity epidemic started during late 70's when the obesity rates shot up from 12-14 % to about 22-25 % (New York Times). He also adds that starting 1977, government started telling Americans to eat less fat. Since then a variety of diets such as low-fat, low-carb etc have hit the market. In general, awareness of obesity is growing. From October to December 1999 there were fewer than 50 articles in the American press about obesity and overweight, whereas during October to December 2002, there were more than 1200 such articles (Wellness Junction). Thus we find it only natural to study the obesity scenario and compare two relatively recent years. In this paper, we investigate the factors that affect Body Mass Index (BMI) and obesity. We believe that body weight is a function of both individual characteristics and external factors, such as state unemployment rate, food insecurity and residence location. To this effect we adopt a multilevel approach. In the next section we explain the data that has been used to conduct the analyses. We then discuss the methodology and empirical results. In the final section we present the conclusions, and discuss scope for future research.

Data

Our goal is to examine several individual level and state level socio-economic factors that might explain the rising phenomenon of obesity in the United States. To address this we use individual level data from the Behavioral Risk Factor Surveillance System (BRFSS), and state level data from the Economic Censuses (EC), Consumer Expenditure Surveys (CES), Economic Research Service (ERS) and Bureau of Labor Statistics (BLS). We study two relatively recent time periods with sufficient gap in between to investigate the changes in the factors contributing to rising BMI and obesity, if at all. Particularly the years 1997 and 2002 are chosen because EC are conducted every five years, ending with 2s and 7s.

The BRFSS was established in 1984 by the Centers for Disease Control and Prevention (CDC). It conducts telephone surveys annually to monitor state level prevalence of major behavioral risks among adults associated with premature morbidity and mortality, which are useful for planning, initiating, supporting, and evaluating health promotion and disease prevention programs. From BRFSS, we obtain the individual level data; specifically, age, education, gender, have kids or not, income, marital status, race, work status, self-reported health status, smoking status, participation in physical activity, consumption of fruits and vegetables, and most importantly the BMI. These surveys interview individuals who are18 years of age or older only. Interviewers ask the height and weight of respondents, and then calculate the BMI themselves.

BMI is calculated as the ratio of weight in kilograms to the square of height in meters. At the individual level, we retain only those respondents who provide complete information on the demographics and other variables of interest. Also, we discard information on respondents who are 95 years and older, since their BMI prove to be outliers more often. Specifically, 31 such observations are deleted from year 1997 and 44 observations from year 2002. After these considerations, we have 104,519 observations in year 1997 and 190,982 observations in year 2002. BMI of 30 kg/m^2 or higher implies obesity. Thus, obesity is derived from BMI.

State level data including annual average state unemployment rates and percentage of the state population living in metropolitan statistical areas are obtained from the BLS. Economic Censuses (EC) gather information on industrial and business activities, and include Census of Retail Trade, Census of Wholesale Trade, etc. EC provides us information on sales of full-service restaurants, fast-food restaurants and grocery stores for each state. We convert these values into per capita sales by dividing the sales by the population estimates of each state. Information on Arizona and District of Columbia are masked in 2002. Thus, we analyze all 50 states and District of Columbia in 1997, whereas in 2002 we analyze only 49 states.

The full-service industry comprises of establishments primarily engaged in (1) providing food services where patrons generally order or select items and pay after eating, or (2) selling a specialty snack or nonalcoholic beverage for consumption on or near the premises. Food and drink may be consumed on the premises, taken out, or delivered to the customer's location. Some establishments (except snack and nonalcoholic beverage bars) in this industry may provide these food services in combination with selling alcoholic beverages. The fast-food industry comprises of establishments primarily engaged in providing food services (except snack and nonalcoholic beverage bars) where patrons generally order or select items and pay before eating. Food and drink may be consumed on premises, taken out, or delivered to customers' location. Some establishments in this industry may provide these food services in combination with selling alcoholic beverages.

From CES, which are surveys conducted by the Bureau of Labor Statistics (BLS), we get data on household's average annual expenditures ($) on food-at-home (FAH), food-away-from-home (FAFH) and other-food-at-home (OFAH). Since 1984, CES has been conducted every year. These surveys collect data on household income and socioeconomic characteristics, and may be used to conduct economic research, market research studies, construction of statistical measures of consumption, etc. The difference between FAH and OFAH is that the former consists of cereals and bakery products, meats, poultry, fish and eggs, dairy products, and, fruits and vegetables, whereas the latter consists of sugar and other sweets, fats and oils, miscellaneous foods, and, nonalcoholic beverages. FAFH is the term used to describe all food prepared outside the home, including food prepared and eaten at restaurants and fast-food establishments, take-out meals prepared by restaurants and fast-food establishments, ready-to-eat meals from supermarkets, and home-delivered meals.

One hurdle at this point is that CES provides regional data on average annual expenditures on FAFH, FAH and OFAH. We convert the regional data into state data by assigning the same value to each state under a particular region. From preliminary analysis using regional data as level 3 variables, i.e. individuals nested within states which in turn are nested within regions, we do not find significant clustering of BMI within regions. We discuss parts of this important preliminary analysis later in 'Analyses and Results' section. Thus, we hope that assigning the same regional values to many states under a specific region will not introduce large biases. In the Appendix (A1) we provide the grouping of states within regions as defined by the U.S. Census Bureau.

The per capita sales of full-service restaurants, fast-food restaurants and grocery stores serve as proxies for availability of various kinds of food, whereas the average annual expenditures on FAFH, FAH and OFAH serve as proxies for food intakes. The ERS is the main source of economic information and research from the U.S. Department of Agriculture. ERS provides us information on prevalence rates of food insecurity with or without hunger in each state.

Finally, for year 1997, two additional state variables are used – percentage of people not exercising sufficiently and percentage of people not consuming adequate amounts of fruits and vegetables. Both these factors are available at the individual level itself in 2002 from BRFSS. Also, in 2002, an additional race category was used – multiracial non-Hispanic. Table 1 gives the estimates and standard deviations for the variables of interest for the two years.

Table 1:
Descriptive statistics

	1997		2002	
Variables	Mean	SD	Mean	SD
State Level				
Avg. annual expenditure ($) per household on				
FAH	2841.26	165.66	3125.62	193.92
OFAH	885.97	62.03	986.45	49.78
FAFH	1961.02	151.03	2255.72	166.53
Per capita sales ($) of				
Full-service restaurants	423.70	124.18	499.83	116.69
Fast-food restaurants	350.50	58.28	395.99	67.51
Grocery stores	1389.64	186.41	1439.12	293.52
Percentage food insecure (with or without hunger)	10.67	2.22	11.09	2.56
Unemployment rate (%)	4.72	1.16	5.36	0.96
Percentage in metropolitan area	69.39	21.53	71.98	21.11
Percentage not exercising enough	27.93	6.93	NA	NA
Percentage not eating enough fruits & vegetables	76.13	4.46	NA	NA
Individual Level				
(continuous variables)				
Age	45.77	16.76	47.74	16.65
BMI	25.91	5.05	26.80	5.49
(categorical variables)	Percentage		Percentage	
Obese	17.3		22.5	
Education				
College or higher	27.8		32.2	

Some college	28.7	27.4
HS or lower	43.6	40.4
Gender		
Male	43.3	42.5
Female	56.7	57.5
Children		
No child	60.9	62.0
At least one child	39.1	38.0
Race		
White non-Hispanic	82.5	81.4
Black non-Hispanic	8.2	7.3
Other non-Hispanic	3.9	4.3
Hispanic	5.4	5.5
Multiracial non-Hispanic (2002)	NA	1.5
Marital		
Never been married	16.6	15.8
Divorced/widowed/separated	25.7	26.5
Married/unmarried couple	57.7	57.8
Work		
Employed for wages	57.7	54.7
Self-employed	8.9	9.4
Unemployed		4.2
Unable to work	3.3	4.5
Retired/homemaker/student	26.8	27.2
Income		
$50,000 and above	26.7	43.9
$20,000 through $49,999	50.3	36.6
$19,999 and less	23.0	19.5
Smoke		
Currently smokes	23.8	22.8
Former smoker	24.4	26.4
Never smoked	51.8	50.7
Self-reported health		
Excellent/very good/good	86.7	84.5
Fair/poor	13.3	15.5
Health-plan		
Has health insurance	87.7	87.6
Does not have health insurance	12.3	12.4
Fruits and vegetables (only in 2002)		
Less than 3 servings/day	NA	38.4
3 but less than 5 servings/day	NA	37.4
5 servings/day or more	NA	24.1
Exercise-outside work (only in 2002)		
Exercises regularly	NA	76.5
Exercises irregularly or never	NA	23.5

We conduct the analysis for two different response variables. The first is the continuous variable BMI. The second response variable is binary that decides whether a respondent is obese or not. As obesity is derived from BMI itself, we present the results for obesity only in the Appendix (A2), whereas the results for BMI are embedded into the document. We will discuss the specifics of the estimation in more details in the next section.

Methodology

Primarily, we are interested in investigating the factors that explain rising BMI in the United States, and if these factors contribute differently during the two years – 1997 and 2002, under study. In addition, we believe that

BMI is not only a function of individual behavior, but is also a result of external environment. Thus we want to model the following:

$$\text{BMI} = f(\underset{\substack{\uparrow \\ \text{individual-level}}}{\text{individual behavior}} + \underset{\substack{\uparrow \\ \text{state-level}}}{\text{environmental factors}})$$

To this extent, we use hierarchical or multilevel modeling (Raudenbush and Bryk) with two levels for our analyses since our datasets consist of individuals nested within states. For the binary dependent variable – obesity, we conduct a multilevel logistic regression. We incorporate the state level information into the individual level by constructing random intercept hierarchical models for each of the above mentioned response variables.

Random intercept hierarchical model may be represented as follows:

Level 1 model: $$y_{ij} = \beta_{0j} + \beta X_{ij} + \varepsilon_{ij} \quad (1)$$

with, $\varepsilon_{ij} \sim N(0, \sigma^2)$

Level 2 model: $$\beta_{0j} = \alpha_{00} + \alpha Z_{0j} + \mu_{0j} \quad (2)$$

with, $\mu_{0j} \sim N(0, \tau^2)$

where,

y_{ij} is the observed BMI of i^{th} respondent in the j^{th} state;

β_{0j} is the intercept in the level 1 model;

β is the vector of parameters for corresponding level 1 sociodemographic characteristics given by the vector X_{ij};

ε_{ij} is the random component in level 1 model;

α_{00} is the intercept in the level 2 model;

α is the vector of parameters for corresponding level 2 state characteristics given by the vector Z_{0j};

μ_{0j} is the random component in level 2 model;

Thus, the combined model, which is analyzed, may be written as follows:

For continuous response variable: $$y_{ij} = [\alpha_{00} + \alpha Z_{0j} + \beta X_{ij}] + [\mu_{0j} + \varepsilon_{ij}] \quad (3)$$

We would like to emphasize that this combined model is a sum of two parts – fixed and random – separated by brackets. The three terms in the first bracket, two alpha terms, and one beta term represent the fixed part. The two terms in the second bracket represent the random part, where μ_{0j} represents the variation in intercepts among states, and ε_{ij} is the variation among individuals nested within states. Also, these two errors terms are independent. We estimate these random effects through their variance components. For the continuous dependent variable BMI, we assume normally distributed error terms at all levels.

Although there are several ways to measure obesity, such as calipers, underwater weighting, computerized topography etc, BMI is the cheapest and easiest way to assess overweight and obesity (CDC). We model obesity as:

$q_i^* = y_i - k_i > 0$

where,

y_i is the BMI of individual *i*;

k_i is a specific threshold for individual *i* and is unobserved;

Thus, the unobserved latent variable q_i^* is such that

$q_{ij}^* = \beta_{0j} + \beta X_{ij} + \varepsilon_{ij}$, for individual in state *j*

Let q_{ij} be an observable version of q_{ij}^*, such that

$$q_{ij} = \begin{cases} 1, & q_{ij}^* > 0 \\ 0, & otherwise \end{cases}$$

Thus, the probability that the i^{th} respondent in the j^{th} state is obese is given by

$P_{ij} = P(q_{ij} = 1) = P(q_{ij}^* > 0) = F(\beta_{0j} + \beta X_{ij})$

The level 2 equation is same as before, in equation (2). It is a linear regression model.
Multilevel logistic regression is conducted for this binary dependent variable. We assume that the error term at individual level, ε_{ij}, has Bernoulli distribution, and μ_{0j}, the error term at state level, is still normally distributed. Hence, the log transformation of the odds of being obese is a logit.

In level 1 model, the BMI of individuals or indication of obesity is expressed as a function of various individual characteristics. We consider the intercept β_{0j} to be a random variable that varies with states, and is expressed as a function of state specific variables Z_{0j}. For simplicity, the slopes at all levels are assumed to be constant. Thus, this model provides a convenient framework for studying multilevel data and systematically analyzes how covariates measured at various levels of a hierarchical structure affect the outcome variable.

If we could get individual level data with all relevant information then standard regression analyses would have sufficed. Efficient estimation and accurate hypothesis testing based on the ordinary least squares (OLS) regression require that the random errors are independent, normally distributed, and have constant variance. Structurally the data is hierarchical because individuals are nested within states, however. There are variables measured on individuals and each state. Because individuals tend to share certain state characteristics, the primary assumption of independence among observations no longer applies, i.e. individuals from a state are more homogenous than if randomly sampled from a larger population. Under the violation of this assumption, OLS regression produces standard errors that are too small. This, in turn, leads to a higher probability of rejection of a null hypothesis (Cohen et al.). Historically, there are three approaches with OLS regression while dealing with hierarchical structure. The first approach is to ignore this structure and give each individual the group or cluster values. Thus, one is indeed fitting the model: $y_{ij} = \mu + \gamma X_{ij} + r_{ij}$

where,

y_{ij} is the BMI of i^{th} individual in j^{th} state;

γ is the vector of parameters for corresponding explanatory variables given by the vector X_{ij};

X_{ij} consists of both state and individual level characteristics;

r_{ij} is the random component;

This approach immediately violates the central assumption of independence (Cohen et al.). The results from such an OLS regression on our hierarchical data and multilevel regression are given in the next section, and we show that OLS does produce smaller standard errors compared to the multilevel model for BMI.

The second approach is to obtain a mean on each predictor variable and the dependent variable for each cluster rather than individual level values. This analysis, also called the aggregate analysis, fails to capture the within group information, leading to inaccurate conclusions (Raudenbush and Bryk). This is because the relations between aggregated variables are much stronger, and can be very different from the relations between the individual level variables.

The third OLS approach is to analyze the regression of the dependent variable on predictors at the individual level, but also include a set of dummy variables to represent the clusters. This method focuses on the relationship of the individual level predictors to the dependent variable when differences among group means are taken care of (Cohen et al.). This is often called the fixed effect approach to clustering, and if the number of clusters is small, then this method is recommended for the analysis of nested data (Snijders and Bosker). This approach is the analysis of covariance (ANCOVA) model.

The multilevel or hierarchical model is a more precise solution to the issues discussed above, since it takes care of the violation of homoscedasticity. In such models, each cluster or group essentially has a different regression model, with its own intercept and slope. They express relationships among variables within a level, and specify how variables in different levels are associated, as they allow for the partitioning of variance into within-group and between-group components (Raudenbush and Bryk).

We use SAS 'Proc Mixed' to fit the multilevel models, along with SAS macro 'GLMMIX' for multilevel logistic regression (Singer). The SAS 'Proc Mixed' codes used in the analysis are documented and explained in the Appendix (A3). 'Proc Mixed' accounts for the effects of clustering by including both level 1 and level 2 predictors. Essentially equation (3) is analyzed such that the intercept from equation (1) is declared as both fixed and random to account for the group structure. Restricted Maximum Likelihood (REML) estimation is the default method in the 'Proc Mixed' procedure. It yields asymptotically efficient estimators for balanced as well as unbalanced designs; this is a clear advantage over ANCOVA methods in modeling real data. The asymptotic normality of REML estimators, furthermore, conveniently allows making inferences on the covariance parameters of the model, which is difficult to do in general linear model.

Analyses and Results

We begin by modeling BMI as a response to several individual level, and state level characteristics for years 1997 and 2002. We want to test the important association between BMI and several state level, and individual level variables. For example, we hypothesize that FAH will have negative association with BMI, whereas FAFH and OFAH will have positive association. Similarly, we expect per capita sales of grocery stores will have negative association, whereas per capita sales of fast-food restaurants will have positive association with BMI. Estimated coefficients and standard errors from the multilevel regression analysis are given in Table 2. The comparison (dropped) categories are left as blanks. We would like to remind the readers that percentages of individuals not exercising enough and not eating enough fruits and vegetables are available at state level in year 1997, whereas the same information is available at the individual level itself in year 2002. Additionally, in 1997 there are four categories of race, whereas in 2002, there are five categories with the addition of multiracial non-Hispanic.

Table 2
Results from Multilevel Regression of BMI

	1997		2002	
Variables	Est.	SE	Est.	SE
Covariance Parameter Estimates				
InterceptState	0.0653*	0.0167	0.0333*	0.0081
Residual	24.5479*	0.1074	27.1913*	0.0891
Intercept	20.6992*	0.1669	21.4513*	0.1364
State level (centered)				
Avg. annual expenditure on				
FAH	–0.0024**	0.0011	–0.0016*	0.0004
OFAH	0.0022	0.0019	0.0028**	0.0013
FAFH	0.0001	0.0008	0.0001	0.0004
Per capita sales of				
Full-service restaurants	–0.0005	0.0004	–0.0006	0.0005
Fast-food restaurants	–0.00003	0.0011	–0.0012	0.0004
Grocery stores	0.0001	0.0003	0.0001	0.0002
Percentage food insecure (with or without hunger)	–0.0386	0.0259	–0.0328	0.0194
Unemployment rate	0.1519*	0.0445	0.1101*	0.0411
Percentage in metropolitan area	–0.0039	0.0021	–0.0054*	0.0018
Percentage not exercising enough	–0.0154	0.0085	NA	NA
Percentage not eating enough fruits & vegetables	0.0207	0.0125	NA	NA
Individual level				
Age	0.1903*	0.0065	0.3114*	0.0051
Age-squared	–0.0015*	0.0001	–0.0030*	0.0001
Education				
College or higher	0.0046	0.0418	–0.8960*	0.0326
Some college	–0.0328	0.0383	0.0086	0.0310
HS or less				
Gender				
Male	0.0323	0.0324	1.1301*	0.0258
Female				
Children				
None	–0.0334	0.0381	–0.0499	0.0300
At least one				
Race				
Hispanic	–0.0552	0.0726	0.3471*	0.0571

Black non-Hispanic	–0.1158	0.0613	1.8445*	0.0507
Other non-Hispanic	0.1775*	0.0850	–0.3920*	0.0626
Multiracial non-Hispanic	NA	NA	0.9995*	0.0998
White non-Hispanic				
Marital				
Never been married	–0.5191*	0.1320	–0.5400*	0.1064
Divorced/Widowed	–0.1690	0.1307	–0.9431*	0.1078
Married/Living as couple				
Age*Marital				
Age*Never been married	0.0149*	0.0033	0.0165*	0.0026
Age*Divorced	0.0042	0.0024	0.0125*	0.0019
Age*Married				
Work				
Employed for wages	–0.0003	0.0469	0.3447*	0.0361
Self-employed	0.0491	0.0664	–0.1502*	0.0505
Unemployed	–0.1578	0.0954	0.5780*	0.0678
Unable to work	–0.1295	0.0973	1.0846*	0.0681
Retired/Homemaker/Student				
Income				
50000 and above	–0.1206**	0.0561	–0.2376*	0.0367
20000-49999	–0.0191	0.0440	–0.6899*	0.0444
19999 and less				
Smoke				
Currently smokes	0.0845**	0.0399	–1.4764*	0.0322
Former smoker	0.0407	0.0392	0.1700*	0.0301
Never smoked				
Fruits				
Less than 3 times/day	NA	NA	0.3805*	0.0329
3-5 times/day	NA	NA	0.2159*	0.0318
5 times/day or more	NA	NA		
Exercise (outside work)				
Exercises regularly	NA	NA	–0.9761*	0.0305
Exercises irregularly or never	NA	NA		
Self-reported health				
Excellent/Very good/Good	–0.0009	0.0501	–1.5440*	0.0377
Fair/Poor				
Health-plan				
Has insurance	0.0241	0.0505	0.2020*	0.0399
Does not have insurance				

*p<0.01, **p<.05

We calculated the variance inflation factors during preliminary analyses and did not find significant multicollinearity among the various variables from the two levels. At the very beginning of the table we present the covariance parameter estimates. These are the estimates for the random effects of the model, namely the lowest-level residual ε_{ij} (by default), and the intercept from level 1. We note that both variance components are significantly different from 0. Thus, these estimates suggest that states do differ in BMI, and there is even greater variation among individuals within states.

We would like to remind the readers that information on FAH, OFAH and FAFH are from regional level which was converted into state level data. This could induce the very same bias we have discussed before, which we are trying to counter through multilevel modeling. In Appendix A4 we produce the covariance parameter estimates from a 3-level hierarchical model with individuals at level 1, states at level 2, and regions at level 3. There are only four regions – Northeast, Midwest, South, and West. Thus, to preserve degrees of freedom, we construct three 3-level hierarchical data with only one predictor at the regional level at one time, no predictor at the state level (except the intercept which is entered by default), and all the individual characteristics at the lowest level. We show that the variance component at level 3 is not at all significant for any of the regional level predictors. Thus, in our case it is justifiable to assign the same regional value to all the states under a specific region.

For both years, FAH is associated with lower BMI, whereas in 2002, OFAH has a positive association. This is not very hard to believe, since FAH is generally healthier, whereas OFAH consisting of chips, candies etc are high in fats and sugars. Surprisingly, neither FAFH nor any of the per capita sales values turn out to be significant factors in either year. This is indeed a significant finding. In spite of all the negative implications surrounding restaurant food, we actually do not find any association between that and BMI for either year. High unemployment rate and high BMI are significantly related. And only for 2002, high percentage of residence in metropolitan areas is associated with lower BMI. This could be because people in towns and cities are more health conscious and frequent their local gyms often.

The difference in the two years is most obvious at the individual level. Whereas in 1997 education plays no role, in 2002 higher education is associated with lower BMI. Also, in 2002 men have significantly higher BMI than women, whereas such distinction can not be made for the earlier time period. Compared to White non-Hispanic, in 2002, Hispanic, Black non-Hispanic and multiracial non-Hispanic have high BMI, whereas other non-Hispanic (mainly Asians) have low BMI. Surprisingly, in 1997, only the other non-Hispanic category turns out to be positively associated with higher BMI. We have to keep in mind that in 1997, other non-Hispanic category includes multiracial non-Hispanic, whereas BRFSS provided a separate category for the latter in 2002. Thus the 1997 result could be confounded. From the 2002 results we do see that multiracial non-Hispanic is a close second to Black non-Hispanic in having a strong positive association with BMI.

We have included two quadratic terms - square of age, and interaction between age and marital status. Age is the only continuous predictor at the lowest level. From preliminary analysis we noticed a U-shaped relationship between age and BMI. Thus, the age-squared term is included to capture the curvature in the association between age and BMI. For both years, age has an inverted U-shape with respect to BMI, i.e. though BMI increases with age, it starts decreasing after a certain point. Also from preliminary analyses we found that not including the interaction between age and marital status resulted in opposite direction of association between BMI and 'never been married'. Thus, we include the interaction to show that being single is associated with lower BMI, but with age this association reverses direction. This was the only interaction causing any change at the lower level. Hence no other interaction term was included in the analysis.

Individual work status is important in year 2002 and not at all in 1997 as an explanatory variable. We club together retired, homemakers and students, because from preliminary analyses these three groups show similar trends with respect to their BMI. This is the base category in work status. Compared to this category, only those who are self-employed have lower BMI. Both employed and unemployed show higher BMI, the unemployed more so. We have unemployment rate at state level too. We do not believe that this caused any bias since the results do not change with the exclusion of either the state level or individual level employment information.

We also find that higher income is associated with lower BMI which reinforces the belief that one has to pay to stay in shape. An undesirable result is that though in 1997 smokers have significantly high BMI; in 2002 this association is reversed. Compared to non-smokers, current smokers have lower BMI, whereas former smokers have higher BMI. We can not explain this reversal of sign. Chen, Yen and Eastwood showed that such a result should be interpreted very carefully due to the endogeneity of smoking. The good news is that in 2002 those who report good health have significantly lower BMI. This is not the case in 1997. Thus, individuals are aware that keeping BMI under check is a step towards better health.

Regular exercise and adequate consumption of fruits and vegetables mean lower BMI. Finally, we note that having a health insurance in 2002, is actually associated with higher BMI. This could mean a greater burden on state and federal budgets as BMI continues to rise. Finkelstein, Fiebelkorn, and Wang found that Medicare and Medicaid pay for at least half of obesity-attributable medical expenses. This means that what would otherwise be a matter of personal choice has become a matter of concern for all taxpayers.

Thus, in 2002, individual level predictors have significant associations with BMI, not so in 1997. This is an important finding. Since in this paper we do not test for endogeneity, we can not discuss causal effects. However, we do know that individuals are continuously reminded of the drawbacks of overweight and obesity through news, magazines, commercials etc. So it is not difficult to believe that people are more aware and health conscious today.

Next, we conduct OLS regression on the same dataset (Table 3). Thus, the observations are no longer clustered. There is just one level – individual level. So, all individuals under a specific state get the same values of per capita sales, unemployment rates, average annual expenditures, percentage of food insecurity, and percentage living in metropolitan areas.

Table 3

Results from OLS Regression of BMI

	1997		2002	
Variables	Est.	SE	Est.	SE
Intercept	20.7197*	0.1628	21.4688*	0.1336
State-level (centered)				
Avg. annual expenditure on				
FAH	–0.0003*	0.0004	–0.0015*	0.0002
OFAH	0.0029*	0.0007	0.0020*	0.0005
FAFH	0.0005	0.0003	0.0004**	0.0002
Per capita sales of				
Full-service restaurants	–0.0006*	0.0002	–0.0008*	0.0002
Fast-food restaurants	0.0001	0.0004	–0.0009*	0.0003
Grocery stores	–0.000002	0.0001	–0.0001	0.0001
Percentage food insecure (with or without hunger)	–0.0333*	0.0107	–0.0346*	0.0084
Unemployment rate	0.1493*	0.0188	0.1016*	0.0180
Percentage in metropolitan area	–0.0031*	0.0009	–0.0043*	0.0001
Percentage not exercising enough	–0.0116*	0.0035	NA	NA
Percentage not eating enough fruits & vegetables	0.0146*	0.0048	NA	NA
Individual-level (not centered)				
Age	0.1902*	0.0065	0.3117*	0.0051
Age-squared	–0.0015*	0.0001	–0.0030*	0.0001
Education				
College or higher	–0.0077	0.0418	–0.9120*	0.0326
Some college	–0.0404	0.0383	0.0019	0.0310
HS or less				
Gender				
Male	0.0351	0.0324	1.1316*	0.0258
Female				
Children				
None	–0.0368	0.0381	–0.0525	0.0301
At least one				
Race				
Hispanic	–0.0805	0.0713	0.3082*	0.0559
Black non-Hispanic	–0.1069	0.0605	1.8171*	0.0501
Other non-Hispanic	0.1697**	0.0829	–0.3852*	0.0613
Multiracial non-Hispanic	NA	NA	1.0257*	0.0991
White non-Hispanic				
Marital				
Never been married	–0.5070*	0.1321	–0.5398*	0.1065
Divorced/Widowed	–0.1553	0.1308	–0.9392*	0.1078
Married/Living as couple				
Age*Marital				
Age*Never been married	0.0148*	0.0034	0.0164*	0.0026
Age*Divorced	0.0040	0.0024	0.0123*	0.0019
Age*Married				
Work				
Employed for wages	0.0021	0.0469	0.3439*	0.0361
Self-employed	0.0529	0.0664	–0.1589*	0.0505
Unemployed	–0.1481	0.0954	0.5773*	0.0678
Unable to work	–0.1247	0.0973	1.0795*	0.0681
Retired/Homemaker/Student				
Income				
50000 and above	–0.1206**	0.0561	–0.2406*	0.0367

20000-49999	–0.0204	0.0440	–0.6975*	0.0443
19999 and less				
Smoke				
Currently smokes	0.0856**	0.0398	–1.4730*	0.0321
Former smoker	0.0434	0.0391	0.1696*	0.0300
Never smoked				
Fruits				
Less than 3 times/day	NA	NA	0.3834*	0.0329
3-5 times/day	NA	NA	0.2139*	0.0318
5 times/day or more	NA	NA		
Exercise (outside work)				
Exercises regularly	NA	NA	–0.9732*	0.0305
Exercises irregularly or never	NA	NA		
Self-reported health				
Excellent/Very good/Good	–0.0048	0.0502	–1.5424*	0.0377
Fair/Poor				
Health-plan				
Has insurance	0.0262	0.0505	0.2062*	0.0399
Does not have insurance				

*p<0.01, **p<.05

There is very little to no difference in either the estimates or standard errors for the individual level variables when OLS regression is used for our hierarchical data. However, there is one significant difference in the state level variables – OLS regression produces smaller variances, although the estimates are not affected much. As discussed before, the immediate consequence of this is that the probability of rejecting the null hypotheses now increases (Cohen et al.). This is clear on comparing Table 2 and 3 results.

First looking at 1997, we see that now OFAH, per capita sales of full service restaurants, percentage food insecure, percentage in metropolitan area, percentage not exercising enough, and percentage not eating enough fruits and vegetables have significant association to BMI using OLS regression. One odd and unexplainable outcome is that percentage not exercising enough seems to have a negative association with BMI. Again for 2002, OLS regression shows significance of additional state variables – FAFH, per capita sales of full-service and fast-food restaurants, and percentage food insecure. Oddly, higher per capita sales from full-service and fast-food restaurants show strong association with lower BMI. The variables that turned out to be significant using random intercept multilevel model are significant here too, and the direction of association is intact. As mentioned before, we are using multilevel data and only the results from multilevel model should be used for interpretation.

Conclusions

We followed a multilevel approach to locate significant explanatory variables for the increasing trend in BMI between years 1997 and 2002. There were two levels under scrutiny – individual, and state. We also conducted OLS regression on the same data, and found that this method produced smaller standard errors for the state level variables compared to multilevel regression, thus increasing the likelihood of rejecting the null hypotheses.

The most significant difference between the two years is that in 2002 almost all individual level characteristics turned out to be significant, not so in 1997. In general, increase in age signifies increase in BMI. However, as is shown by the quadratic term, after a certain age BMI starts to decrease again. Men have significantly higher BMI than women. The only individual level predictor that is insignificant in both years is whether individual has children or not. Individuals with higher education have significantly lower BMI. Among the several race categories, Hispanics, Black non-Hispanics and multiracial non-Hispanics have very high BMI compared to White non-Hispanics, whereas other non-Hispanics have lower BMI. In 1997, only other non-Hispanics had significantly higher BMI than White non-Hispanics. This is not necessarily a major change between the two years; in 1997, BRFSS included multiracial non-Hispanics into the other non-Hispanic category. Among the several marital status categories in 2002, we find that though being single (includes those never been married, divorced and widowed) is associated with low BMI, with increasing age this direction is reversed.

Compared to retirees, homemakers and students, employed individuals have higher BMI, but not as high as those who are unemployed or unable to work. This is not surprising; most jobs are sedentary in developed countries.

Yet, unemployed individuals have more trouble controlling their weights. We also found evidence that higher income is associated with lower BMI. Consumption of adequate amounts of fruits and vegetables and participation in regular physical activities implied lower BMI in 2002. In 1997 both variables were used at the state level as percentages, and did not turn out to be significant. In 2002, people with health-insurance tended to have higher BMI; this should worry the state and federal health budgeters. Finally, in 2002, people who reported excellent, very good or good health, had significantly lower BMI than those who reported fair or poor health, implying increasing awareness.

The one individual level predictor that confuses is the smoking status. In 1997, smokers have higher BMI compared to non-smokers, whereas in 2002, smokers have significantly lower BMI than non-smokers and former smokers have higher BMI. Since this is not a panel data, we can not put forth any explanation for this complete role reversal. However, as mentioned before, previous studies have shown smoking to be an endogenous variable while analyzing BMI.

At the state level, high unemployment rate stands out as being associated with higher BMI for both years. Another variable that is consistently associated with lower BMI is FAH, which is a good sign.

Additional variables such as kind of job (blue collar or white collar), proximity of fast-food restaurants from work place, whether or not parents and/or close relatives are obese, etc would have contributed greatly to this study. For simplicity we used only one type of multilevel structure – random intercept hierarchical model. However, one could also try random slopes and random intercept plus slopes hierarchical models. Such models would test if the characteristics at one particular level are affected by the characteristics from other levels. In spite of these issues we obtained many significant outcomes from this research. As far as we know, this is a first multilevel approach to model individual's BMI. Most of our findings for the individual level are consistent with those of previous studies (Philipson; Chou et al.). What is different and new is the insignificance of 'important' state variables such average annual expenditure per household on FAFH, and per capita sales of full-service and fast-food restaurants in their association with BMI. And this is true for both years. Almost all individual level variables have significant association with BMI in 2002, which is not the case in 1997. We think this is an important finding considering 2002 is a recent time period and hence its results are more relevant today.

More individuals today recognize high BMI as a health hazard. Those who exercise, consume a healthy amount of fruits and vegetables, and depend more on food cooked at home, are suitably fit and healthy. However, we know that certain groups of people are more susceptible, such as Hispanics, Black non-Hispanics, multiracial non-Hispanics, individuals who are unemployed or are unable to work, individuals from lower income categories, and less educated people. They need immediate attention given that this epidemic has been around for a while now.

Appendix

A1

US Census Bureau Regions with States

Region 1: Northeast		
Connecticut		New Jersey
Maine		New York
Massachusetts		Pennsylvania
New Hampshire		
Rhode Island		
Vermont		
Region 2: Midwest		
Indiana		Iowa
Illinois		Kansas
Michigan		Minnesota
Ohio		Missouri
Wisconsin		Nebraska
		North Dakota
		South Dakota
Region 3: South		
Delaware	Alabama	Arkansas

District of Columbia	Kentucky	Louisiana
Florida	Mississippi	Oklahoma
Georgia	Tennessee	Texas
Maryland		
North Carolina		
South Carolina		
Virginia		
West Virginia		
Region 4: West		
Arizona		Alaska
Colorado		California
Idaho		Hawaii
New Mexico		Oregon
Montana		Washington
Utah		
Nevada		
Wyoming		

A2

Results from Multilevel Logistic Regression of Obesity

	1997		2002	
Variables	Est.	SE	Est.	SE
Covariance Parameter Estimates				
InterceptState	0.0139*	0.0041	0.0059*	0.0017
Residual	0.9978*	0.0043	0.9993*	0.0033
Intercept	–3.3685*	0.0947	–2.8515*	0.0684
State level (centered)				
Avg. annual expenditure on				
FAH	–0.0007	0.0005	–0.0005*	0.0002
OFAH	0.0002	0.0009	0.0009	0.0001
FAFH	–0.0002	0.0004	0.0001	0.0002
Per capita sales of				
Full-service restaurants	–0.0004**	0.0002	–0.0004	0.0002
Fast-food restaurants	0.0002	0.0005	–0.0004	0.0003
Grocery stores	–0.00003	0.0001	–0.00001	0.0001
Percentage food insecure (with or without hunger)	–0.0207	0.0123	–0.0144	0.0083
Unemployment rate	0.0649*	0.0212	0.0315	0.0177
Percentage in metropolitan area	–0.0015	0.0010	–0.0015	0.0008
Percentage not exercising enough	–0.0073	0.0041	NA	NA
Percentage not eating enough fruits & vegetables	0.0086	0.0059	NA	NA
Individual level				
Age	0.0662*	0.0037	0.1109*	0.0026
Age-squared	–0.0005*	0.00004	–0.0011*	0.00003
Education				
College or higher	0.0062	0.0225	–0.3558*	0.0158
Some college	–0.0014	0.0207	0.0007	0.0143
HS or less				
Gender				
Male	0.0257	0.0174	0.0960*	0.0122
Female				
Children				
None	–0.0251	0.0210	–0.0168	0.0143
At least one				

Race					
	Hispanic	–0.0250	0.0412	0.0356*	0.0267
	Black non-Hispanic	–0.0456	0.0331	0.5484*	0.0216
	Other non-Hispanic	0.1263*	0.0458	–0.1054*	0.0311
	Multiracial non-Hispanic	NA	NA	0.3463*	0.0443
	White non-Hispanic				
Marital					
	Never been married	–0.1918*	0.0767	–0.1910*	0.0543
	Divorced/Widowed	–0.1078	0.0724	–0.3853*	0.0540
	Married/Living as couple				
Age*Marital					
	Age*Never been married	0.0057*	0.0018	0.0062*	0.0013
	Age*Divorced	0.0025	0.0013	0.0053*	0.0010
	Age*Married				
Work					
	Employed for wages	–0.0223	0.0252	0.0994*	0.0174
	Self-employed	0.0028	0.0351	–0.0878*	0.0246
	Unemployed	0.0009	0.0523	0.1866*	0.0309
	Unable to work	–0.0758	0.0504	0.2071*	0.0289
	Retired/Homemaker/Student				
Income					
	50000 and above	–0.0139	0.0304	–0.1324*	0.0169
	20000-49999	–0.0082	0.0237	–0.3160*	0.0210
	19999 and less				
Smoke					
	Currently smokes	0.0435**	0.0215	–0.5181*	0.0158
	Former smoker	0.0148	0.006	0.0806*	0.0139
	Never smoked				
Fruits					
	Less than 3 times/day	NA	NA	0.1610*	0.0157
	3-5 times/day	NA	NA	0.0806*	0.0155
	5 times/day or more	NA	NA		
Exercise (outside work)					
	Exercises regularly	NA	NA	–0.4042*	0.0135
	Exercises irregularly or never	NA	NA		
Self-reported health					
	Excellent/Very good/Good	–0.0127	0.0259	–0.5656*	0.0163
	Fair/Poor				
Health-plan					
	Has insurance	0.0024	0.0280	0.0764*	0.0186
	Does not have insurance				

*p<0.01, **p<.05

A3

SAS codes for Proc Mixed

```
proc mixed data covtest convh=1E-2;
class state <level 1 variables>;
model bmi = <level 1 variables> <level 2 variables> /solution ddfm=bw;
random intercept/subject=state;
run:
```

The COVTEST option on the PROC MIXED statement produces the hypothesis tests for the variance and covariance components. The CLASS statement declares the categorical variables. The MODEL statement is used to indicate the fixed effects and the RANDOM statement for random effects. BMI on the left-hand side of the MODEL statement indicates the dependent variable. The intercept is entered by default in the MODEL statement. By writing ‘intercept’ in the RANDOM statement, we declare that the intercept should also be treated as a random effect. SOLUTION asks SAS to print the estimates for the fixed effects. DDFM=BW tells SAS to use the between-within method for computing the denominator degrees of freedom for the fixed effects. The SUBJECT option specifies the multilevel structure, indicating how level 1 variables are clustered into level 2 variables. Here, the subgroups are designated by the classification variable STATE. Without this option, the variance component representing the state effect would be omitted.

A4
Covariance parameter estimates from 3-level hierarchical models

Level-3 Predictor	Random effects		1997 Estimate (SE)	1997 p-value	2002 Estimate (SE)	2002 p-value
FAH	Intercept	Region	0.0166 (0.0237)	0.2417	0.0153 (0.0210)	0.2336
	Intercept	State (Region)	0.0845 (0.0181)	<0.0001	0.0755 (0.0169)	<0.0001
	Residual		24.5490 (0.1074)	<0.0001	27.1519 (0.0879)	<0.0001
OFAH	Intercept	Region	0.0896 (0.0809)	0.1340	0.0560 (0.0586)	0.1697
	Intercept	State (Region)	0.0817 (0.0171)	<0.0001	0.0697 (0.0148)	<0.0001
	Residual		24.5495 (0.1074)	<0.0001	27.1525 (0.0879)	<0.0001
FAFH	Intercept	Region	0.0150 (0.0198)	0.2241	0.0508 (0.0572)	0.1873
	Intercept	State (Region)	0.0805 (0.0167)	<0.0001	0.0747 (0.0166)	<0.0001
	Residual		24.5496 (0.1074)	<0.0001	27.1520 (0.0879)	<0.0001

References

American Obesity Association. http://www.obesity.org/

Cassell, J.A., (1995). Social Anthropology and Nutrition: A Different Look at Obesity in America. Journal of the American Dietetic Association, 424-427.

Centers for Disease Control and Prevention (CDC). http://www.cdc.gov/

Chen, Z., Yen, S., and Eastwood, D., (2005). Does Cigarette Consumption Have a Causal effect on Weight Reduction? Submitted to Journal of Family Economics and Issues in April 2005.

Chou, S., Grossman, M., and Saffer, H., (2004). An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System. Journal of Health Economics, 23, 565-587.

Cohen, J., Cohen, P., West, S.G., and Aiken, L.S., (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (Third Edition). Mahwah, NJ: Earlbaum.

Guthrie, J.F., Lin, B.H., and Frazão, E., (2002). Role of Food Prepared Away from Home in the American Diet, 1997-78 versus 1994-96: Changes and Consequences. Journal of Nutrition Education and Behavior, 34, 11-16.

Kuchler, F., and Golan, E., (2004). Is There a Role for Government in Reducing the Prevalence of Overweight and Obesity? Choices, an online magazine of the American Agricultural Economics Association, 19(3rd quarter).

McGinnis, J.M., and Foege, W.H., (1993). Actual Causes of Death in the United States. Journal of the American Medical Association, 270, 2207-2212.

Philipson, T., (2001). The World-wide Growth in Obesity: An Economic Research Agenda. Health Economics, 10, 1-7.

Raudenbush, S.W., and Bryk, A.S., (2002). Hierarchical Linear Models: Applications and Data Analysis Methods (Second Edition). Newbury Park, CA: Sage.

Saad, L., (2003). Public Balks at Obesity Lawsuits. Princeton, NJ: The Gallup Organization.

Singer, J.D., (1998). Using SAS Proc Mixed to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. Journal of Educational and Behavioral Statistics, 24, 323-355.

Snijders, T.A.B., and Bosker, R.J., (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks, CA: Sage.

Taubes, G., (2002). What if it’s all been a Big Fat Lie? New York Times Magazine. Issue July 7, 2002.

Wellness Junction. http://www.wellnessjunction.com/athome/selfcare/epidemic.htm

Endnotes

[1]Doctoral Candidate, Department of Agricultural, Environmental and Development Economics, 2120 Fyffe Road, Columbus, OH 43210, Ph: 614-292-8339, Fax: 614-292-7710, mandal.7@osu.edu

[2]Professor, Department of Agricultural, Environmental and Development Economics, 2120 Fyffe Road, Columbus, OH 43210, Ph: 614-292-6414, Fax: 614-292-7710, chern.1@osu.edu