ABSTRACT

KASSENS, ALICE LOUISE. Do The Obese Respond to Adverse Health Events?: A Bayesian and Behavioral Approach (Under the direction of Dr. Alvin Headen)

Obesity is an epidemic that is particularly prevalent amongst aging Americans. The economic literature concerning this phenomenon is minimal, and largely focuses on its causes. This dissertation examines the issue from a different direction. First, a Bayesian model is developed to determine how aging Americans use obesity related health information, and if their response differs from the non-obese. Secondly, a behavioral model is employed to see if the same health information elicits a behavioral response. This is the first behavioral response model to determine the existence and magnitude of BMI defined behavioral changes after exposure to new personal obesity related risk information. The estimates of the first model suggest that the obese are Bayesian updaters to a certain extent, and that they utilize adverse health information differently than the non-obese. The behavioral estimates imply that the obese adjust their weight after an adverse health diagnosis. Several suggestions for further work and extensions of these models are given.

Do the Obese Respond to Adverse Health Events? : A Bayesian and Behavioral Approach

By
Alice Louise Kassens

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

ECONOMICS

Raleigh

2005

APPROVED BY:

_____ _____

_____ _____
Chair of the Advisory Committee

DEDICATION

I dedicate this dissertation to Dr. William Rodgers. As my undergraduate economics professor and advisor at the College of William and Mary, he supported me and fostered self-confidence in my abilities as no other teacher has done before. I would never have thought graduate school possible without his direction, encouragement, and instruction. If my only achievement as a professor of economics is to inspire someone as he did me, I will have had a very successful career indeed.

BIOGRAPHY

Alice Louise Kassens is the daughter of Drs. William and Catherine Kassens. She grew up in Wilmington, NC, and attended the Groton School from 1988 to 1993. Alice went to the College of William and Mary, where she earned a B.A. in economics and history. In 1998 she entered the doctoral program for economics at North Carolina State University, and graduated in 2005. Her fields of interest include health and labor economics.

Alice is a professional runner in her spare time. She began running in college and competed in cross-country and track for the College of William and Mary, and was a member of several NCAA Division I Academic All-American teams. Alice continued running after college, and signed a professional contract with Brooks in April 2004. Marathons are her focus, and she hopes to qualify for the U.S. Olympic Trials in that event.

Alice's parents, Bill and Katie, are physicians in Wilmington, NC. Alice has a younger sister, Catherine (Kate) Helen Kassens, who is a computer programmer in Wilmington, NC. Kate attended the University of North Carolina at Chapel Hill, and the sisters are life-long Tarheel basketball fans. Alice has three dogs, and three cats, Dean, Kennedy, Millie, Hillary, Gabby, and Szabo, respectively.

ACKNOWLEDGEMENTS

I thank my parents, professors, and dissertation committee for their support, patience, and encouragement throughout my graduate school experience. I express gratitude to my writing partners, Teensy and Millie Kassens. Finally, I appreciate the patience and tolerance Dean, Kennedy, Hillary, Gabby, and Szabo displayed during this process, and I promise to be a more available friend from now on.

Table of Contents

List of Tables

I. Introduction

The prevalence of obesity in the United States has risen dramatically in recent decades. It has been estimated that between 1976 and 2000 the percentage of obese individuals has more than doubled, from 15% to 31% of the population[1]. This gain is greatest among older adults. In 2000, an astonishing 36% of men and 40% of women between the ages of 60 and 74 were considered obese, a 22% and 18% increase, respectively, since 1976. This trend covers all race, age, income and educational groups. The growth in obesity is especially alarming as epidemiological studies show an increase in associated mortality. In fact, individuals who are considered to be obese have a 50 to 100 percent increased risk of premature death compared to the non-obese[2]. It has been estimated that 300,000 deaths a year may be attributed to obesity[3].

Obesity has only recently been addressed in the economic literature. These works attempt to explain the growth in obesity rates through the actions of individuals, and focus on factors including prices and time inconsistency. This paper takes a different path towards understanding the prevalence of obesity. Instead of focusing on why individuals are overweight and obese, this work assumes the state of obesity to be an optimal choice rendered from a utility maximizing model. The goals of this dissertation are two-fold. First, using a Bayesian framework, this paper examines how the obese process information when forming their subjective longevity perceptions. Secondly, a

---

[1] Flegal et al. "Prevalence and Trends in Obesity Among US Adults, 1999-2000." *JAMA*, 2002, Volume 288, No. 14, pp. 1723-1729.
[2] NIH, NHLBI. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. HHS, PHS; 1998, p. 23.
[3] Sturm, R, KB Wells. "Does obesity contribute as much to morbidity as poverty or smoking?" *Public Health* 2001 May; 115(3):229-35.

behavioral model is developed to determine how the new information affects obese behavior.

This paper is the first effort to empirically evaluate how new information, in the form of health shocks, affects various BMI groups' longevity perceptions. Smith et al. attempted to answer the same question with regards to smokers and non-smokers[4]. Their model is replicated to determine what type of updating rules BMI subpopulations use to revise their longevity assessment. Smith et al. selected the Bayesian model to yield insight into the manner smokers use information. They suggested that smokers are more likely to be risk takers, and thus may have different propensities to acquire health risk information. Smokers are said to fail to associate information they do attain with the risky activities they undertake. Under this setup, Smith et al, hypothesized that current smokers ignore some information, and weight the information they do use when forming their longevity perceptions differently than non-smokers. It is this author's contention that overeating is also one such risky activity, and that the obese are just as likely to ignore health risk information, and consider it differently as smokers. Therefore, the Bayesian model is also applicable here. The behavioral model is an obvious extension of the Bayesian model to observe if the longevity perceptions are manifested in a behavioral response.

The RAND-Health and Retirement Survey (HRS) Version B data set is supplemented with the raw HRS The Health and Retirement Survey (HRS) for the analysis. The HRS is appropriate for this study for two reasons. It provides the needed

---

[4] Smith, Taylor, Sloan, Johnson and Desvousages. "Do Smokers Respond to Health Shocks?" *The Review of Economics and Statistics*, November 2001, 83(4): 675-687.

information in a panel form, which allows for a robust and thorough analysis compared to a single cross section, and it permits an investigation of the older population that is experiencing the greatest increase in obesity rates. The health shocks contained in this dataset can be categorized as obesity related and general. Knowing how the obese respond to these different types of shocks, compared to the normal group, can help to frame an effective public health policy.

The remainder of this paper is organized as follows: Section II is a review of the literature. Section III defines obesity and describes the trends in the United States. Section IV outlines the conceptual model, and is followed by Section V and the Empirical Strategy. Section VI describes the data used in the analysis. The Bayesian model is presented in Section VII, and the behavioral model is introduced in Section VIII. These sections include the model specifications, estimation, and results. This dissertation closes in Section IX with the conclusion, which includes a critique of the work and suggestions for further research. References and appendices follow.

II. Literature Review

It is no longer solely "the business of a comic poet to paint the vices and follies of human kind[5]," but also that of economists. Excesses in eating, drinking, gambling and the decision to smoke or use illegal drugs, do not fit the traditional rational economic framework and were ignored in economic literature for some time. Rationality is the cornerstone of microeconomic theory. The assumption of rationality requires utility maximizing individuals to account for the future consequences of their actions when they are making choices today. Excesses, such as the ones afore mentioned, seem to be the antithesis of rational behavior. Microeconomists sought a way to incorporate this "irrational" behavior into their theory of consumer choice so they too could "paint the vices and follies of human kind".

The goal of this literature review is twofold. First, a thorough review of the Smith et al. paper is given. This is followed by a discussion of the seminal implementation of the Bayesian updating framework to explain individual decisions and behavior set forth by Viscusi. This review points out the criticisms of the Bayesian framework, and discusses the argument that the risk adjustment process implied by the model fits within the guise of rational decision-making. Secondly, the current economic models used to explain excessive eating and the increase in obesity rates are reviewed. Much of the recent literature has focused on vices such as smoking, alcohol and illegal drugs, and has only recently incorporated the phenomenon of obesity. Thus, an analysis of the economics of obesity, such as this paper, must include the theories that have been tested

---

[5] William Congreve, *The Double Dealer*, epistle dedicatory.

solely on other risky behaviors, as these theories have framed the thought behind the models used to address obesity.

a. Informational Response:

The increased research in obesity, including that by economists described in this literature review, is reflective of the mounting concern of the upward trend in obesity rates. To date, the economic analyses have focused on the causes of obesity and the consumption patterns of the obese. There has been no research addressing how obese individuals respond to different types of information and how that new information is transferred to their behavior. That is the goal of this paper. Individuals participate in risky behavior of multiple forms, including smoking, speeding, and overeating. If these individuals are rational, they are engaging in this behavior as an optimal choice, given all of the available information. The risks associated with many events are not known, due to lack of experimental data, and a subjective measure of these risks must be formed.

There have been several attempts in the literature to evaluate individual risk perception and its impact on behavior. Viscusi (1985) based his work on the assumption that individuals incorporate new information consistently when forming probabilistic judgments. For example, he said, workers learn from job searches and consumers learn about the products that they prefer. This learning process suggests that past beliefs, along with new information, are used in the creation of present beliefs. Viscusi used his Bayesian updating model to examine smokers' behavior (1990, 1991, 1992, 1995). He

found that the higher an individual's risk perception is from dying from lung cancer, the less likely it is for the individual to smoke[6].

Smith et al. advanced Viscusi's model and offered greater detail regarding both the risk perceptions of smokers versus non-smokers and the significance of prior beliefs on the posterior report. They examined how smokers use new information acquired through health shocks relative to non-smokers and former-smokers. Their model is the basis of this dissertation, and is used to investigate how obese individuals respond to personal health shocks relative to normal individuals.

i. Smith, Taylor, Sloan, Johnson and Desvousges:

Smith et al investigate how smokers adjust their personal longevity expectations after exposure to new health information, compared to non-smokers and former smokers in "Do Smokers Respond to Health Shocks?[7]." They use the initial two waves of panel data from the Health and Retirement Study (HRS) for their analysis. The dataset consists of a representative national sample of birth cohorts aged 51-61 in 1992, and their spouses if married. Baseline interviews are conducted in the respondent's home and are followed up with telephone interviews in subsequent years. HRS includes a wide variety of information on health behaviors, health and functional status, and the individual's subjective perception of longevity. Smith et al. use the individual's expected probability of living to age 75 as the dependent variable to measure the respondent's reaction to new

---

[6] Additional support for this finding can be found in Viscusi et al. (2000), and Liu and Hsieh (1995).
[7] Smith, V. Kerry, Donald H. Taylor, Jr., Frank A. Sloan, F. Reed Johnson, and William H. Desvousges. "Do Smokers Respond to Health Shocks?" *The Review of Economics and Statistics,* November 2001, 83(4): 675-687.

information. The new information is in the guise of a health shock, and is categorized as either a general or smoking related event.

Smith et al. use the Bayesian updating framework to model respondents' subjective perceptions of longevity and smoking risks. Viscusi[8] first proposed this model, but it has since been employed in several applications. According to the Bayesian updating model, a prior probability, or baseline, is used, in addition to any other information received, to form the current, or posterior probability. In this study, the prior is the subjective probability of living to age 75 in the initial wave of data, and the subsequent subjective probability is considered the posterior probability report. The authors first conduct chi-square tests of two hypotheses, before attempting several estimates of the Bayesian model. The first hypothesis tests if smokers, former smokers and never smokers have different longevity perceptions, while the second investigates if these groups react differently to unanticipated, exogenous health shocks related to smoking in comparison to those shocks that are not smoking related.

The dependent variable is the respondent's self reported subjective probability of living to 75 years of age[9]. This question is asked in each wave, and the authors rescaled the responses to the zero-one scale to better represent probabilities. The explanatory variables include a lag of the dependent variable, demographic measures (age, education, gender, race, marital status, whether parents alive), current smoking status (at the time of

---

[8] Viscusi, W. Kip and O'Connor. "Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?" *American Economic Review,* 74(5), 942-956. Viscusi, W. Kip. "A Bayesian Perspective on Biases in Risk Perceptions," *Economics Letters* 17:1 (1985), 59-62.

[9] Because these variables are subjective, they are prone to reporting error, and their reliability is questionable. Several researchers have conducted tests to determine if these responses are legitimate estimates that can be used in regression analysis such as this. I return to this issue in the data section of my analysis, as the HRS is the dataset employed to conduct my tests as well.

the interview), indicators of cognitive ability, and three health measures. The three health measures include the general (GS) and smoking related (SS) health shocks, changes in preexisting conditions (ΔPC), and changes in functional ability (ΔAD). All health shocks are events that occurred between the two waves, and represent the onset of a new, serious, and unexpected health condition. The smoking related health shocks include cardiovascular and cerebrovascular diseases, smoking related cancers, and chronic lung disease. General health shocks include serious medical conditions not related to smoking, including diabetes, accidents knocking the respondent unconscious or requiring hospitalization, surgery to replace a joint after a fracture, non-smoking related cancers, and any other condition requiring at least a three day hospital stay.

Smith et al.'s examination begins with two simple hypothesis tests. The first hypothesis tests if the longevity expectations in wave 2 for the respondents in the two waves are comparable across smoking groups. A chi-square test of the distribution comparability rejects the null hypothesis for all three smoking groups in both waves. The second hypothesis suggests that the three smoking groups react differently to health shocks. This chi-square approach imposes fewer restrictions than the formal Bayesian model used later, and indicates that current smokers only respond to smoking related health shocks, while the other two groups react to both types of shocks. Respondents react to the information by lowering their subjective longevity expectation.

Differences in characteristics between smoking groups may also affect the way respondents react to health shocks, thus Smith et al. formalize their Bayesian model via ordinary least squares. This method controls for a wider range of potential contributing factors than the simple chi square tests. In this model, perceived longevity in wave two,

P_t, is a weighted function of a respondent's initial longevity perception, P_{t-1}, and an unobserved risk equivalent, r_t. The risk equivalent is based on any new information an individual receives that initiates a longevity revision. Respondents attach a precision to both the baseline longevity perception, θ, and the risk equivalent, γ. The posterior assessment of living to age 75 in wave 2 is then:

$$P_t = \frac{\theta P_{t-1} + \gamma r_t}{\theta + \gamma} \quad (1)$$

This is the same posterior assessment that Viscusi and O'Connor assigned to workers' perceived risk of radon in the workplace[10]. Each of the three types of health information is included in P_t, in addition to respondent characteristics. New information can come in the form of changes in preexisting conditions, specific health shocks experienced between the two waves, and changes in self-reported ability to undertake specific activities. If no change in a preexisting condition is reported, this is viewed as good information, and is hypothesized to have a positive effect on longevity perception. Deteriorations in health are likewise hypothesized to reduce the longevity report. Using the Bayesian terminology, P_t is the posterior probability and P_{t-1} is the prior probability.

The unobservable risk equivalent of new information, r_t, is modeled as a function of the health indicators:

$$r_t = f(SS_{t-1}, GS_{t-1}, \Delta PC_t, \Delta AR_t) \quad (2)$$

Demographic and socioeconomic variables, including age, gender, race, and education, are incorporated into the updating framework to control for individual methods of

[10] Viscusi and O'Connor (1984).

utilizing new health information. A vector of such variables, labeled z_k, is added to r_t, and when (2) is substituted into (1), the smokers risk perception becomes:

$$P_t = \frac{\theta}{\theta+\gamma} * P_{t-1} + \frac{\gamma}{\theta+\gamma} * f(SS_{t-1}, GS_{t-1}, \Delta PC_t, \Delta AR_t, z_1, z_2, ..., z_k) \quad (3)$$

Although this formulation is similar to Viscusi's model (1984, 1990), there is an important difference. Viscusi only had a single cross section, and therefore was not able to observe the prior subjective belief. The only source of updating in his model was the difference in the information available to the various demographic groups. Viscusi's estimating equation is similar to:

$$P_t = \beta_0 + \sum_{j=1}^{k} \beta_j z_j \quad (4a)$$

Under his formulation, the effects of P_{t-1} and r_t are captured by the other determinants of P_t, namely the z_j's. Smith et al. have two waves of data at their disposal, and their estimating equation is:

$$P_t = \frac{\theta}{\theta+\gamma} * P_{t-1} + \alpha_0 + \sum_{j=1}^{k} \alpha_j x_j \quad (4b)$$

where the health measures, SS_{t-1}, GS_{t-1}, ΔPC_t, ΔAR_t, and the demographic variables are combined in x_j in equation 4b. The two waves of data allow for the evolution of the subjective beliefs to be measured over time, and the role of $\frac{\theta}{\theta+\gamma}$ can be distinguished from α_j. All of the estimates use the Heckman two-step selection framework to adjust for the respondents who did not answer the longevity question. The authors also estimated the equation via OLS and ordered probit models, but the conclusions remain the same as those rendered under the Heckman method. The estimates from the updating model show

that smoking related health shocks are the only forms of new information current smokers consider when revising their longevity perceptions. The other two smoking groups are also influenced by general health shocks. These results support the findings of the chi square tests discussed earlier.

Why do current smokers react differently to various health shocks than never smokers and former smokers? The authors suggest that current smokers are more likely to be "risk takers," and may have different propensities to acquire information, which they fail to associate with risky activities they undertake. Smoking can be considered one of these risky activities. If this is true, current smokers ignore some information, and weight the information they do use when forming their longevity perceptions differently than non-smokers. Smith et al. conduct two tests to compare the risk equivalents of three smoking groups, the first testing if the risk equivalents of smoking related health shocks across groups are equal, and the second comparing the average risk equivalent of new health information for each group.

The smoking related health shock coefficients were normalized by one minus the weight assigned to the prior longevity perception, $(1-\theta)$, in the first test. This ratio is computed for each group. The results indicate that current smokers assign a greater weight to the smoking related health shocks than former smokers, but the difference between current smokers and never smokers is less significant. The second test utilizes a reduced form expression for the risk equivalent of the new information for each individual:

$$r_t = \frac{(\alpha_0 + \sum_{j=1}^{k} \alpha_j x_{jt})}{1 - \left(\frac{\theta}{\theta + \gamma}\right)} \quad (5)$$

The matrix compares each sub-sample's estimated updating model to the observed experience of all three sub-samples when calculating r_t. This estimation procedure allows for the separation of the effects of each group's weighted information and their actual health experiences. For example, the first column suggest that if current smokers experience the health outcomes of the never smokers, they still report a lower subjective probability of living to age 75. This indicates that the risk equivalent of the new health information assigned by current smokers is smaller than what never smokers would assign to their own health experiences.

In all cases, smokers consistently differ from never smokers and former smokers in how they use new information when forming their personal longevity expectations. Smith et al. concluded that current smokers are more optimistic than their behavior warrants prior a health shock, but once an adverse, smoking related health condition occurs, current smokers update their expectation more dramatically than the other groups. This suggests current smokers attach a larger risk equivalent to these shocks. They do not react comparably to general health shocks.

Can the Bayesian model yield insight into the thought processes of the obese? More specifically, do obese individuals use new information to update their subjective probability of living to age 75 differently than the non-obese? One goal of this dissertation is to answer this question. To do so, the Smith et al. framework is used, but instead of comparing smokers to never smokers and former smokers, obese individuals

are compared to normal individuals, as defined by reported BMI. This model is appropriate, as the obese are more likely to be considered "risk takers," and may also have different propensities to acquire new information than the non-obese. Being obese can be considered a risky state of being, just as Smith et al. posited about the smokers compared to the other smoking groups. Thus, the obese could ignore some information, and weight the information that they do use when forming their subjective longevity perception differently than the non-obese.

b. Historical Use of the Bayesian Framework in the Economic Literature

Researchers in fields including economics and psychology have studied individual risk perceptions and behavior under uncertainty over the years, but the results are inconclusive[11]. Some of this research suggests that economic agents are behaving irrationally when presented with uncertainties. A common example is that given by Kunreuther (1976). He documents seemingly risk adverse consumers failing to purchase heavily subsidized federal flood insurance. Kip Viscusi is the most prevalent economist in the area of behavior under uncertainty, and it is his work that Smith et al. base their aforementioned model upon. He maintains in all of his papers that individuals behave rationally when making choices under uncertainty. The framework in which he commonly characterizes the decision making process of these individuals is a Bayesian

[11] Some of these works include Fischoff et al. (1983, 1993), Schoenbaum (1997), Smith et al. (1987, 1988, 1990, 1995), Svenson (1984), Taylor et al. (1984, 2000), Viscusi (1985, 1990, 1992), Viscusi, et al. (1984, 1987, 1998).

updating model. Viscusi begins several of his papers with a version of one of the following statements that define this work:

> *A fundamental issue in the economics of uncertainty is how individuals process information and make choices under uncertainty*[12].
>
> And
>
> *A fundamental aspect of risk perception in the economics literature is the ability of individuals to incorporate new information reliably in forming their probabilistic judgments*[13].

Viscusi and O'Connor (1984) published the first study to investigate the dynamics between risk assessments and individual decisions. More formally, they sought to analyze the characteristics of chemical workers' risk assessments about their jobs, how the workers used new information about the level of risk at their job, and how this information impacted individual decisions. Viscusi and O'Connor assumed that a Bayesian process, in which their subjective beliefs follow a beta distribution, characterized the workers' learning process. They modeled the posterior probability, p_i, of an adverse job outcome after receiving a hazard warning for chemical i as:

$$p_i = \frac{\gamma p + \xi_i s_i}{\gamma + \xi_i} = \frac{\xi_i s_i}{\gamma + \xi_i} + \frac{\gamma p}{\gamma + \xi_i} \quad (6)$$

[12] Viscusi and O'Connor (1984)
[13] Viscusi (1984)

where p is the prior subjective belief, s is the new information given in the form of new risk information about chemical i, and the coefficients are the associated relative weights or precisions. The corresponding regressive equation used for each chemical i is:

$$RISK1_i = \alpha_i + \beta_i RISK_i + \mu_i \quad (7)$$

in which

$$\alpha_i = \frac{\xi_i s_i}{\gamma + \xi_i}, \beta_i = \frac{\gamma}{\gamma + \xi_i} \quad (8)$$

The estimated parameters, $\hat{\alpha}_i$ and $\hat{\beta}_i$, are used to construct two key measures of the information conveyed in the chemical warnings. The risk level s_i is defined as

$$s_i = -\frac{\alpha_i}{(\beta_i - 1)} \quad (9)$$

and the informational content of the new chemical warning relative to the prior information, p, is

$$\psi_i \equiv \frac{\xi_i}{\gamma} = \left(\frac{1}{\beta_i}\right) - 1 \quad (10)$$

The higher the value of ψ_i, the greater the weight the worker places on the new warning label compared to their prior judgment.

Viscusi and O'Connor found that after a given hazard warning concerning a chemical used by a worker, employees revised their risk assessment in the expected direction: a new reported lower risk conveyed by the label resulted in a reduction in the risk assessment on the part of the worker. The prior belief also impacted the posterior report, making the posterior report a weighted function of both the prior risk assessment and the new information presented to the worker. These results support the idea that workers are Bayesian updaters and that although they initially may choose a sub-optimal

outcome, they use new information to eventually arrive at the optimum. Behavior under uncertainty in this case is best characterized as falling somewhere between the optimal decision given perfect prior information and random irrational decision-making.

Smoking behavior is used as a case study to test the Bayesian updating model of risk perceptions and the resulting potentially risky consumption decisions in Viscusi's 1991 paper, "Age Variations in Risk Perceptions and Smoking Decisions[14]." As the title implies, the influence of smoking risk behavior is tested for various age cohorts. The dataset used in the examination is extensive, but Viscusi notes that its cross sectional nature prohibits the investigation of the role of information over time, and the impact of new information on changes in smoking behavior. In addition, the only adverse smoking event included is lung cancer. His results indicate that individual perceptions of smoking risks influence smoking behavior. The youngest age group revealed the highest risk perception and was more likely to overestimate the risk than the population at large. This result is to be expected since this age cohort receives a large share of its smoking information from recent, publicly provided sources, which often lead to overestimations of health risks.

---

[14] Viscusi, Kip W. "Age Variations in Risk Perceptions and Smoking Decisions," *The Review of Economics and Statistics*, Volume 73, No. 4 (Nov., 1991), 577-588.

c. Informational Response and Obesity: Kamhon Kan and Wei-Der Tsai[15]

Several economists have investigated health risk knowledge and its influence on individual behavior with regards to smoking, drinking, and other risky activities. Kan and Tsai are the first to do so for the obese. Obesity health risk is defined as individual awareness of the harmful health consequences that obese individuals are likely to incur due to their unhealthy state. Kan and Tsai model obesity as a choice made by the individual based upon the level of knowledge held by him or her. An individual who is aware of the relationship between obesity and its health risks will associate higher costs with obesity than an individual with lesser knowledge. Therefore, the possession of knowledge concerning obesity health risks prevents an individual from being obese.

The framework imposed by Kan and Tsai implicitly assumes individual rationality. Becoming obese is a decision made based upon an internal cost-benefit analysis, and is assumed to be the optimal choice for the individual given the level of health risk knowledge in his or her possession. This dissertation uses the utility maximizing choice framework as well. Utility maximizing models differ from other models of obesity presented later in this literature review. Those models present obesity from a physiological and genetic standpoint.

The empirical work uses data from the *Cardiovascular Disease Risk Factors Two-Township Study* in Taiwan. It is a longitudinal survey that examines the relationship between the risk factors of cardiovascular disease and the development of the diseases.

---

[15] Kan, Kamhon and Wei-Der Tsai, "Obesity and Risk Knowledge." *Journal of Health Economics*, 23 (2004), 907-934.

The health risks were determined by asking respondents if they think that obesity will cause:

(1) Apoplexy
(2) Hypertension
(3) Diabetes
(4) Heart disease
(5) Gout
(6) Breast cancer
(7) Ulcer

The possible responses were:

(A) Very likely
(B) Possible
(C) Not possible
(D) Do not know

Scores of 3 for (A), 2 for (B), 1 for (C), and 0 for (D) were assigned to construct a variable measuring each respondent's risk knowledge concerning obesity for each of the seven adverse health events included in the study. The final dataset consisted of 3,700 respondents, including 1,729 males and 1,974 females.

The quantile regression technique is used in the estimation process to estimate the effects of the included covariates on the quantiles of the BMI distribution. Kan and Tsai suggest that this method is well suited for a study of obesity since the impact of the explanatory variables on the entire BMI distribution can be estimated. Alternative methods include least squares and least squares with an indicator variable for obesity. Least squares is not deemed appropriate since there is an optimal range for individual BMI, and BMI is not a monotonic indicator of health. Including an indicator for obesity

is an improvement on simple least squares, but this method is also problematic. The cutoff values for being normal, overweight, and obese are arbitrary and referential, and obesity related health risks effect individuals falling just to the left of the obese classification.

Individual i's BMI is denoted as B_i, and the quantile regression model for the θth quantile is shown as:

$$B_i = \beta'_\theta X_i + \varepsilon_{\theta i} \quad (11)$$

$$Q_\theta(B_i \mid X_i) = \beta'_\theta X_i \quad (12)$$

where β_θ is a vector of coefficients, X_i is a vector of demographic characteristics, $\varepsilon_{\theta i}$ is a stochastic term, $Q_\theta(B_i \mid X_i)$ represents the conditional quantile of B_i given X_i, θ is an index for the quantile, and i is an index for the respondents. The coefficients are estimated by solving the minimand

$$\min_{\beta_\theta} \sum_i Q_\theta(B_i - \beta'_\theta X_i \mid X_i) \quad (13)$$

The θth conditional quantile of B_i given X_i is shown as $Q_\theta(B_i \mid X_i) = \beta'_\theta X_i$. The kth element of vector β_θ refers to the marginal effect of the kth covariate,

$$\beta_{\theta k} = \frac{\partial Q_\theta(B_i \mid X_i)}{\partial X_{ik}} \quad (14)$$

This can be interpreted as the marginal change in the θth conditional quantile as a result of a change in X_{ik}. This marginal effect may vary over different quantiles. The coefficients are estimated over nineteen quantiles for a given set of regressors.

The empirical strategy is a two-stage procedure, with an ordered probit model in the first stage. The estimation is conducted for males and females separately since the

authors hypothesize that males and females may respond differently to health risk information. Denoting a particular health risk knowledge variable for individual i as H_{ki}, where k=1,…,7, the model can be written as:

$$H_{ki} = j, \text{ if } \mu_{kj} < H_{ki}^{*} \leq \mu_{k,j+1}, \text{ j} = 1,\ldots,4; \quad (15)$$

$$H_{ki}^{*} = \delta_{k}' Z_{i} + u_{ki}; \quad (16)$$

$$\mu_{k0} = -\infty, \mu_{k4} = \infty; \quad (17)$$

$$u_{ki} \sim N(0,1); \quad (18)$$

where μ_{kj} and δ_{k} are the parameters to be estimated, Z_i is a vector of individual characteristics, and the random variable u_{ki} is allowed to be correlated with the residual of the BMI equation. H_{ki}^{*} represents a continuous latent measure of individual i's perceived likelihood for the kth adverse health outcome to occur.

The estimation results indicate that males, with a BMI around or below the median, increase their BMI with additional health risk information. The authors suggest that this is possibly due to the fact that these men, who are knowledgeable about the obesity health risks, are also likely familiar with the aspects of good nutrition in general. With the exception of the uppermost percentile, the health risk coefficient is insignificant at the upper end of the distribution. At the 95th percentile, however, where BMI is equal to 29.41, an increase in health risk knowledge by one-unit results in a reduction in BMI of 1.48 units. In sum, mildly overweight men are unresponsive to additional health risk knowledge, while men close to categorical obesity heed increased health risk warnings. The mildly overweight men may not believe that the adverse health events associated with obesity will affect them since they are not by definition obese.

The response of females to additional health risk information across the BMI distribution is markedly different from their male counterparts. The health risk coefficient is consistently close to zero, and statistically insignificant at all percentiles. Females do not seem to alter their BMI, even at the farthest end of the distribution, when presented with additional information concerning the harmful, and even deadly, effects of obesity.

The Kan and Tsai model suggests that further research into the behavioral response of individuals to obesity health risk information needs to control for gender, as there are significant differences in the interpretation of such knowledge between males and females. The trends in obesity vary more across age groups than gender, as does the incidence of obesity related health events. In the United States, individuals between the ages of 50 and 69 exhibit the highest growth rates in obesity for both men and women over the period 1960 and 2000.[16] These age groups, along with their old peers also experience greater incidence of obesity related health shocks. Thus, it is equally compelling to stratify the dataset to account for these differences when investigating the response to knowledge of the risks associated with obesity, as it is to partition the analysis according to gender. Smoking related health events are also more frequent with age, and this is why Smith et al. selected the HRS dataset for their paper. This dissertation focuses on the aging American population, and the HRS is well suited for this purpose.

Respondents in the Kan and Tsai dataset gather health risk information from the media. Other controls for informational exposure include the respondent's frequency of

[16] Flegal et al. (2002).

meetings with friends, education level, and participation in community activities. These three measures attempt to capture "word of mouth" resources. It well known that individuals respond to personal experiences, whether they be direct or indirect. There is no measure in Kan and Tsai's model to indicate if the responses personally experienced the included adverse health events. First hand information may have a profound influence on individual health risk perceptions. Additionally, previous beliefs, as mentioned in the review of the Bayesian literature, have been shown to impact present risk perceptions, and perhaps current behavior stemming from those beliefs. This suggests an additional improvement on the Kan and Tsai model is to include indicators of prior risk knowledge, when attempting to ascertain the influence of current knowledge on current BMI. It is conceivable that an individual recently exposed to new information will display a lagged response. It may take time for a change in BMI to take affect after exposure to health risk information. Based on these criticisms, this dissertation seeks to improve upon the infant literature on obesity, and the corresponding health risk perceptions, with a Bayesian updating model.

d. Risky Behavior: An Economic Explanation

Most of the literature to date regarding obesity has focused upon finding an explanation for its prevalence and growth in the United States and abroad. Why are people fat? Eating is a necessity of life, but in excess it is widely known to be harmful. Still, as a society, our waistlines are expanding at an alarming rate. This seems irrational. Perhaps it is merely a matter of taste. Could it be that some consumers, though rational,

have a greater taste for overeating than others, and it is optimal for them to do so? This was the argument made by Kan and Tsai, and also that of this dissertation. Often, however, once resolved to be a difference in tastes, disputes have often been ended because it is thought that there is no room for rational persuasion. Stigler and Becker[17] state that this conclusion is not a reason to stop the examination of a behavior, and propose that tastes are in fact stable and homogenous for all consumers, and something else is driving some to excess. The differences in behavior are due to something deeper than a mere difference in taste, such as prices or income.

Stigler and Becker considered the commodity "euphoria," which is produced by the consumption of heroin, in their 1977 work, "De Gustibus Non Est Disputandum." The most significant statement made by the authors concerning individual behavior is that the use of heroin increases with exposure, while the amount of euphoria falls if the demand for euphoria is sufficiently inelastic. Stigler and Becker interpreted this to mean that addiction to heroin, a growth in use with exposure, is the result of the inelasticity of demand for the commodity, not the cause of the inelasticity. Their conclusion suggests that the elasticity can be used to determine if a good is beneficially or harmfully addictive. A high elasticity suggests a beneficial addiction, while a low elasticity implies a harmful addiction.

Tastes need not change to explain why the use of heroin increases with exposure, or why addicts are insensitive to price changes. If tastes are constant, Stigler and Becker showed that use grows with exposure and heroin is addictive because demand is inelastic,

---

[17] Becker, Gary S., and George J. Stigler. "De Gustibus Non Est Disputandum". *The American Economic Review*, Volume 67, Issue 2 (March 1977), 76-90.

not vice versa. A policy implication from these results is that an increase in taxes, or more severe punishments, will likely have little influence on the use of harmfully addictive goods, and a large impact on beneficially addictive goods. Stigler and Becker successfully reconciled additive behavior with stable preferences, and showed that their hypothesis produced useful predictions about observable behavior, without abandoning previously characterized irrational behavior, including heroin addiction.

e. Rational Addiction:

Stigler and Becker's paper was the first to stir the pot on the thought of excessive, often unhealthy, behavior displayed by many consumers. It served as a springboard for the vast literature on habit formation and addiction, and many works built upon Stigler and Becker's brief, but insightful work[18]. Psychologists have studied addictions for years, and their clinical findings are incorporated into economists' models of such behavior. Three important characteristics were observed in experimental studies that helped define addictive behavior: reinforcement, tolerance and withdrawal. Reinforcement indicates that greater consumption today of a good leads to greater consumption of the good in the future. Tolerance implies that given levels of consumption today are less satisfying when past consumption has been greater. Withdrawal is the negative physical reaction and other losses in satisfaction that is experienced when consumption is ceased.

---

[18] Michael Grossman gives a brief, yet informative, overview of the addiction literature in Ch. 10 *The Economics of Human Behavior.*

Gary Becker, in collaboration with Kevin Murphy, authored the seminal work in the rational addiction literature, "A Theory of Rational Addiction[19]" in 1988. Under the rational addiction structure, consumers are farsighted in the sense that they anticipate the consequences of their current actions. The consumption of a good is said to be addictive if an increase in present consumption results in an increase in future consumption, a definition reminiscent of Stigler and Becker's earlier designation: an increase in use with exposure. Thus, present consumption is positively related to future consumption. This framework can be used to model the consumption pattern of the obese compared to the non-obese, although to date, the only attempt is that by Cawley in his 1999 dissertation work[20].

"A Theory of Rational Addiction" initiated a host papers that sought to explain addictions to various goods under the guise of rationality[21]. Many of these papers left the Becker and Murphy model unchanged, and used consumption and price data to find evidence of, or lack there of, rational addiction of their chosen vice. Becker, Grossman and Murphy's 1994 publication, "An Empirical Analysis of Cigarette Addiction[22]," tests the Becker and Murphy model of rational addiction by looking at the reaction of cigarette consumption to a change in cigarette prices. Explicitly, the authors assess whether lower past and future prices for cigarettes increase current cigarette consumption. The

---

[19] Becker, Gary S. and Kevin M. Murphy. "A Theory of Rational Addiction." *The Journal of Political Economy*, Volume 96, Issue 4 (August 1998), 675-700.
[20] Cawley, John. "Obesity and Addiction." PhD Dissertation, Department of Economics, University of Chicago, 1999.
[21] These works include Winston (1980); Iannaccone (1986); Michaels (1988); Chaloupka (1991); Becker, Grossman, and Murphy (1994); Orphanides and Zervos (1995); Levy-Livermore (1998); Cawley (1999); Coppejans, Mrkaic, and Seig (2000); Gruber and Koszegi (2001); Kenkel, Reed, and Wang (2002); Coppenjans and Seig (2002).
[22] Becker, Gary S., Michael Grossman, and Kevin M. Murphy. "An Empirical Analysis of Cigarette Addiction." *The American Economic Review,* Volume 84, Issue 3 (June, 1994), 396-418.

empirical results support the theory of addictive behavior, as cross price effects are negative and long run responses exceed short run responses.

Frank Chaloupka[23] tested for rational addiction in cigarette consumption under the Becker and Murphy framework, but with noteworthy differences from the Becker, Grossman and Murphy attempt. In "Rational Addictive Behavior and Cigarette Smoking," Chaloupka used data from the National Health and Nutritional Examination Study (NHANES), yielding a microeconomic analysis of the nature of cigarette consumption. Becker et al. used state level data, and made the assumption that the aggregate data accurately reflected individual behavior. Chaloupka did not have to make this assumption. NHANES is a national survey of approximately 28,000 individuals aged 6 months to 74 years, and was conducted between 1976 and 1980. Data was collected on current consumption, lagged consumption, and consumption at the time that the individual smoked his or her greatest average quantity of cigarettes. The survey also included questions concerning the number of years since the individual began smoking regularly, and for former smokers, the number of years that the individual has not smoked. While this micro-level data is at risk of reporting errors, among other issues, it offers an alternative, and perhaps more acceptable, test for rational addiction and cigarette consumption than the Becker et al. paper. In all but one of the estimated equations the coefficients all conformed to the predictions of the rational addiction theory. That is, current consumption was significantly and negatively related to the current price of cigarettes, while past and future prices generally had a positive effect on

---

[23] Chaloupka, Frank. "Rational Active Behavior and Cigarette Smoking." *The Journal of Political Economy*, Volume 99, Issue 4, (August 1991), 722-742.

current consumption. In the equations that include past and future consumption, the coefficient on past consumption was larger in magnitude than that of future consumption. The same was true for the magnitudes of the lagged and lead coefficients of price.

Most tests of the Becker and Murphy model of rational addiction support its predictions. This cannot be said of Jonathan Gruber and Botond Koszegi, who, in their work "Is Addiction 'Rational'? Theory and Evidence[24]," found that past results are fragile. The authors were critical of the Becker and Murphy assumption that individuals are correctly forecasting prices far in advance, due to the fact that few price increases are in fact announced. Also, Gruber and Koszegi noted that the correlation between the lead of price and current consumption is usually considered a failure of a specification test of the model outside the rational addiction model. It is not typically considered evidence of forward looking or rational behavior. Finally, the authors took issue with Becker and Murphy's assumption of individual time consistency. This assumption implies that an individual's future behavior coincides with his or her current desires regarding this behavior.

Gruber and Koszegi suggested that the rational addiction model should consider time inconsistent behavior. They sited psychological evidence that documents such inconsistency. In such experiments, consumers systematically revealed a lower discount rate when making decisions over time intervals that are further away than ones that are closer to the present. While there is no non-experimental evidence of time inconsistent behavior, there is likewise no psychological evidence that consumers are time consistent

---

[24] Gruber, Jonathan and Botond Koszegi. "Is Addiction 'Rational'? Theory and Evidence." NBER Working Paper 7507, January 2000.

either. Gruber and Koszegi's goal was to integrate a more "realistic" approach of intertemporal choice into the Becker and Murphy model. Instead of a time consistent agent who discounts utility at time t using the following:

$$\sum_{j=0}^{T-1} \delta^i U_{t+j} \quad (19)$$

a time inconsistent agent is a quasi-hyperbolic discounter, and discounts utility according to:

$$U_t + \beta \sum_{i=1}^{T-t} \delta^i U_{t+i} \quad (20)$$

where β and δ are assumed to be between zero and one. Gruber and Koszegi were unable to develop and implement a test capable of distinguishing quasi-hyperbolic and exponential preferences, but they successfully introduced a decisive alternative to the traditional Becker and Murphy model of addictive behavior.

f. Hyperbolic Discounters:

Gruber and Koszegi's suggestion that time inconsistency is a better description of addicted agents provides a direct path to the current obesity literature. Few economists have attempted to explain excessive eating as a rational addiction, but there have been several attempts to explain the behavior by alternative methods. One such explanation also incorporates the time inconsistent agent described by Gruber and Koszegi. The fact that some individuals have self-control problems results in gluttony. These self-control problems can be exacerbated by environmental changes, including price changes. Cutler, Glaeser and Shapiro addressed this issue, and provide a link between the actions taken by

smokers compared to non-smokers and those taken by the obese compared to the non-obese.

In their 2003 working paper, "Why Have American's Become More Obese?[25]," Cutler, Glaeser and Shapiro examined the impact of declining food prices on time inconsistent agent's consumption behavior. More explicitly, they attributed the increase in obesity to the reaction of time inconsistent consumers to the reduced time costs of food stemming from technological innovations. Mass production has evolved over the past decades, making consumption of many foods, most notably potatoes, easier[26]. Price reductions are typically viewed as beneficial, but if the consumer is time inconsistent, and thus has self-control problems, the increased consumption may become excessive, resulting in an increase in obesity.

Weight is ultimately determined by net caloric intake over time. An individual gains weight if more calories are consumed than expended. Cutler et al. found that while there has not been a significant change in energy expenditures since the 1980s, there has been an increase in energy intake. This increase is explained via a theory based on the division of labor. Prices and incomes could cause an increase in caloric intake: As people become wealthier, or if the relative price of food declines, they will demand more food. Because the CPI for food items has increased 3 percent slower than that for non-food

---

[25] Cutler, David M., Edward L. Glaeser, and Jesse M. Shapiro. "Why Have Americans Become More Obese?" *National Bureau of Economic Research Working Paper*, January 2003.

[26] Prior to World War II, Americans ate large amounts of potatoes, mainly baked, boiled or mashed, and were usually consumed at home. French fries were rare because of high time costs of preparation. In the post-war period, many innovations centralized French fry production, resulting in a substantial increase in potato consumption both in and outside of the home, mostly in the fried form.

items, and real incomes did not increase greatly for the bottom of the income distribution where obesity has increased, another explanation for increased intake must be found.

Technological innovation has resulted in a reduction in the time cost of food production, which can increase food intake through several venues: 1) increased variety of foods consumed, 2) increased frequency of food consumption, 3) a switch to high calorie/high flavor prepared foods which had previously been unavailable, or 4) an increase in the overall consumption of each individual food item. These responses are stem from the standard price mechanism, and may explain a good portion of the increase in consumption, but are not likely the sole cause. Self-control issues are likely a significant influence as well.

Cutler et al. also employed the hyperbolic discounting framework of Laibson[27] and Harris and Laibson.[28] An individual receives utility from consumption of a durable composite commodity (C) and food, which is represented by caloric intake, K. Utility is lost from being overweight, and to eliminate income effects, utility is assumed to be linear in C. The utility function is then:

$$U_t = C_t + U(K_t) - h * Weight_t \quad (21)$$

The price of food is P, which includes both time and money costs. If Y is income, then $C_t = Y - PK_t$. Following Laibson, individuals discount the future in two ways. The first is the standard exponential utility form, in which a consumer receives $e^{-\rho t}$ units of utility for t periods of time in the future. Second, individuals make a distinction between the

---

[27] Laibson, David. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 1997, Volume 62, p. 443-477.
[28] Harris, Christopher and David Laibson. "Dynamic Choices of Hyperbolic Discounters." *Econometrica*, 2001, Volume 69, Number 4, p. 935-957.

"current self" and the "future self," which is the essence of the hyperbolic model. The future self's utility is discounted by γ, which falls on the interval of $0 \leq \gamma \leq 1$. If $\gamma = 1$, the future is discounted at the same rate as today, but if $\gamma = 0$, the future is ignored completely. People switch from being "current" selves to "future" selves with a hazard rate of λ. Individuals are assumed to have perfect knowledge about all the parameters of the system, but do not know when the switch to the "future" self will occur. This makes the value of future consumption probabilistic, where the future utility remains connected to the current self with probability $e^{-\lambda t}$, and to the future self with probability $1- e^{-\lambda t}$, and is worth only γ as much.

Food consumption carries two costs: the dollar value of forgone consumption and the health and social costs of increased weight. If the composite commodity is assumed to be durable, consumption evolves according to

$$\frac{dC_t}{dt} = -\delta C_t + I_t \quad (22)$$

where $I_t = Y - PK_t$. The marginal value of the forgone consumption of an additional unit of food is then estimated by $\frac{P\gamma}{\rho + \delta}$. Health costs are linear with slope h, and weight evolves according to the differential equation

$$\frac{dW_t}{dt} = -\alpha - \mu W_t + wK_t \quad (23)$$

The health costs of current consumption are approximated by $\frac{wh\gamma}{\rho + \mu}$. In equilibrium, the consumer will choose K such that

$$\frac{P\gamma}{\rho + \delta} + \frac{wh\gamma}{\rho + \mu} = (e^{-\rho t}(e^{-\lambda t} + (1 - e^{-\lambda t})\gamma)U'(K) \quad (24)$$

The equilibrium condition has several implications for consumption and weight. First, the level of food consumption will increase with both decreases in γ and increases in λ. Reductions in γ yield less forward-looking consumers, resulting in more present-oriented consumption. Increases in λ indicate that the value of consumption is less likely to benefit the future self, and leads to increased current consumption. Second, a reduction in the price of food leads to increased food consumption and higher steady state weights. Technological innovation in mass food production affects consumption in two ways: 1) the price P and 2) the delay before consumption λ. The weight gain due to a reduction in time delay is more pronounced for hyperbolic consumers because people with self-control problems are more responsive to the readily available food than those without self-control issues. This theory not only helps to explain the increase in obesity over the past decades, but also the dramatic increase in the upper tail of the weight distribution.

g. Technological Change:

Cutler et al attribute the rise in obesity in the United States to an interaction between individuals with little self-control and technological change. Technological change lowers the time cost of food preparation, which increases food consumption. For individuals with self-control problems, this can result in excessive eating and obesity. In contrast, technological change has also been used to explain the increase in obesity by way of a reduction in energy expenditure, rather than leading to an increase in food

consumption. Philipson and Posner[29] were the first to test this assertion. Their model remains one focused on individual consumption, but unlike Cutler et al, Philipson and Posner directed this focus to examining why individuals exercise less, instead of why are they eating more.

Philipson and Posner state that in developed countries, obesity has grown with modest increases in caloric consumption and a substantial increase in both dieting and exercise. Therein lies the rub. Individuals gain weight when energy consumption exceeds expenditure, but these economists find that the increase in obesity rates observed in recent decades has occurred despite the energy balance apparently shifting towards a reduction in obesity. Philipson and Posner examine this trend in their theoretical paper "The Long Run Growth in Obesity as a Function of Technological Change," and attempted to find an explanation for the recent rise in obesity.

Philipson and Posner and Cutler et al. noted that technological change has lowered the cost of consuming calories, but each approached this observation in a different manner. Cutler et al. theorized that this effect is due to improvements in mass production, which lower the time cost of food preparation. Philipson and Posner argued that technological change has both lowered the time cost of consuming calories and increased the cost of expending calories. This has contributed to the rise in obesity by lowering the real price of food and the physical expenditure of calories per hour in market and household production.

---

[29] Philipson, Tomas J., and Richard A. Posner. "The Long Run Growth in Obesity as a Function of Technological Change." National Bureau of Economic Research Working Paper 7423, 1999.

Technological change decreases the real price of calories because food prices decline at the same time the amount of physical exertion required while working falls. In an agricultural society, work is very strenuous, and the worker is in essence paid to exercise. In addition, the cost of not working can be starvation due to a meager welfare system. In contrast, work in an industrial society requires little physical exertion, and is relatively sedentary. Not working does not necessarily lower weight in this system, as food stamps and other welfare benefits are available. Thus, individuals must pay to exercise in an industrial society, usually in the form of forgone leisure.

Philipson and Posner's interpretation of the effects of technological change offer an explanation for why obesity has increased despite the simultaneous increase in dieting and exercise with a moderate increase in calorie consumption. Technological innovation has resulted in a reduction in on-the-job exercise, so despite the increase in recreational exercise, net caloric expenditure has fallen. If calorie consumption increased, all be it minimally, while energy expenditure fell, the result is an increase in the obesity rate.

Philipson and Posner also hypothesized that the growth in obesity due to technological innovation may be self-limiting. Many authors, beginning with Michael Grossman[30], have found that the relationship between income and weight is concave: as incomes increase for low income individuals, weight increases, but eventually a desire for thinness prevails, and weight begins to fall with further increases in incomes. Weight will not increase indefinitely with income as economies become more advanced, and

---

[30] Grossman, Michael. 1972. *The Demand for Health*: *A Theoretical and Empirical Investigation*. New York: Columbia University Press.

obesity growth due to technological advances becomes self-limiting. Eventually weight will begin to fall as individuals attain greater levels of income.

Utility is defined as a function of weight, food consumption, and alternative consumption, C, or U(W(F,S),F,C). Grossman's observation of the non-monotonicity between utility and weight is incorporated here, so that there is an inverted U-shape relationship between the two. W_0 is the ideal weight that an individual does not prefer to deviate from, holding other consumption constant, and is the weight that would be chosen if there are no costs of doing so. However, W_0 does not necessarily correspond to the weight that maximizes health or longevity, and is most likely not the most economically preferred. If physical activity is held fixed, the most preferred weight, in an economic sense, depends upon opportunities and preferences:

$$MaxU(W(F,S),F,C)$$

$$\text{s.t. } C + pF \leq I \quad (25)$$

where p is the price of food and I is income. The necessary condition for an interior choice of calories is one that balances the weight effect and the joy of eating against the forgone consumption of alternative goods:

$$U_W W_F + U_F = pU_C \quad (26)$$

For many prices and incomes, the preferred weight is different from the ideal weight.

If physical activity is not assumed to be fixed, but is instead a choice variable that is affected by occupation, housing, transportation and exercise, the maximization problem becomes more involved, albeit, more realistic. If I(S) denotes the income that can be earned when S calories are used, then the marginal effect is I_S, which measures the loss, $I_S \leq 0$, or gain, $I_S \geq 0$, in income from a more active lifestyle. It is useful to consider a

full measure of income that includes tradeoffs between higher-earning, but more sedentary jobs, cheaper suburban housing inducing more sedentary forms of transportation, and technological change enabling more efficient, but more sedentary home or market production. The second maximization problem is then:

$$Max_{(F,S)} U(W(S,F),F,I(S)-pF) \quad (27)$$

The necessary conditions for an interior solution are

$$\frac{d}{dF}: U_W W_F + U_F = pU_C \quad (28)$$

$$\frac{d}{dS}: U_W W_S = -I_S U_C \quad (29)$$

The first condition is similar to the previous maximization problem, except now it is also conditional upon the optimal level of calorie expenditure. The second first order condition suggests an individual is overweight if and only if he forgoes income to spend calories, or

$$W \geq W_0 \Leftrightarrow I_S \leq 0. \quad (30)$$

Leisure can be allocated to weight control by way of off-the-job exercise, and the increases in weight due to technological advances can be balanced by increased off-the-job activities. The substitution of leisure-time for work-time physical activity stems from the increased sedentariness of employment. The allocation of time becomes an issue. Consumption of calories requires money income and is thus goods intensive, while physical activity to expend calories is time intensive. Physical activity has to be self-produced, and this takes time. As a result, high wage earners may be more obese because they have more money to buy food and it costs more for them to substitute work for

leisure to exercise. Philipson and Posner incorporated the time allocation decision into their analysis by including both earned and unearned income in the budget constraint:

$$C + pF \geq wH + I \quad (31)$$

where w is the hourly wage rate and H is hours worked. Time is split between work, inactive leisure (L) and active leisure (E):

$$H + L + E \leq T \quad (32)$$

E can also include such weight controlling measures as smoking, taking diet pills and exercise. Total physical activity is defined as

$$S = sH + E \quad (33)$$

where s is calories spent per hour of work.

The individual chooses the number of hours to work and to exercise, as well as the amount of food to consume. The maximization problem becomes:

$$U(W(F, sH + E), I + wH - pF, F, T - H - E) \quad (34)$$

with the following first order conditions for an interior solution:

$$(1)\ U_W W_{SS} + U_C W = U_L \quad \text{(Labor Supply)} \quad (35)$$

$$(2)\ U_W W_S = U_L \quad \text{(Recreational Exercise)} \quad (36)$$

$$(3)\ U_W W_F + U_F = pU_C \quad \text{(Caloric Intake)} \quad (37)$$

Condition (2) implies that only overweight individuals exercise for the purpose of weight control, because for anyone who does exercise, since L is valued and exercise reduces weight, weight must, on the margin, have a negative effect on utility, or else no one would exercise. (1) implies that the utility an exercising person obtains from working an extra hour in a sedentary job is less than for a non-exercising person. Finally, (3) suggests that an exerciser derives more utility, and a non-exerciser less, from the

consumption of other goods, since he is willing to use resources to prevent that consumption from increasing his weight.

Philipson and Posner's hypothesis has implications for public policy. They note that it is important to distinguish between being overweight in a medical sense and being overweight in a social Pareto-inferior way when considering public policy alternatives to combat obesity. Under a rational-choice framework, the former does not demand public intervention, while the later suggests that there are some people who would be willing to pay for others to lose weight. If health is not everything in life, rational people will eat more and exercise less than medical science believes is best. People may prefer high-paying jobs that demand little physical exertion to more physically demanding ones that pay less. By allowing preferences and technology to determine obesity, public interventions that focus on educating the public about the harms of being overweight may have little effect. The issue is not information, the authors conclude, but incentives. Everyone knows how to lose weight: eat less or exercise more. The problem is, few want to pay the price, in effort, expense, or forgone pleasure to maintain a healthy weight.

Darius Lakdawalla and Tomas Philipson empirically tested the Philipson and Posner obesity model in their work "The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination[31]." They first outlined a dynamic theory of weight management that predicted technological change effects both supply- and demand-side forces. The authors stated that the observed relative food price reductions, long-run weight increases, and changes in food consumption can not be explained by

---

[31] Lakdawalla, Darius and Tomas Philipson. "The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination." National Bureau of Economic Research Working Paper 8946, 2002.

technological change on the demand- or supply-side alone. In this model, the supply side innovations work through agricultural changes and the demand side modifications occur via technological innovations that render work more sedentary. Secondly, Lakdawalla and Philipson predicted that earned and unearned income might have differing effects on weight growth. The authors hypothesized that an understanding of these differences could help explain why income is positively related with weight across countries, where technologies vary, but negatively within countries, where technology levels are uniform.

Lakdawalla and Philipson used data from the National Longitudinal Survey of Youth (NLSY) to estimate their model. The dataset begins in 1978 and contains 12,686 individuals aged 14 to 22. The cohort is followed over time, and the last survey included is from 1998. The dataset includes several demographic variables, including detailed income information. The authors choose to only present the results for female workers due to the high level of noise amongst the male results. Lakdawalla and Philipson found support for the charge that technological progress, which has resulted in more sedentary jobs, led to an increase in weight. They further hypothesized that the issue is not solely a demand-side issue, but that technological change also lowers the price of food, creating upward supply-side pressure on weight. Lakdawalla and Philipson's empirical estimates suggest that their model is a successful extension of the Philipson and Posner theory.

Philipson and Posner's theory focused on the technological innovations that accompany a shift from an agricultural society to an industrialized one. They argued that technological change lowered the time cost of consuming calories and increased the cost of expending calories, which contributed to the increase in obesity by lowering the real price of food and the physical expenditure of calories per hour in market and household

production. Because of the increased opportunity cost of non-market and home production, innovations have been made to economize on time previously allocated to these areas. The increase in obesity could be the result of several economic changes that have altered the lifestyle choices of Americans. Chou, Grossman, and Saffer focus on specific societal forces that they hypothesized changed the cost of nutritional and leisure time choices. In particular, "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System"[32] examined the effects of changes in relative prices on individual choices, and is a more intimate analysis of consumer responses, and the resulting increase in obesity rates witnessed in recent times.

The authors selected several economic changes to focus on. First was the increase in the value of time, most notably of women, reflected in the increases in labor force participation rates and the number of hours worked. Because individuals spend more time at work, and less at home, there is a subsequent increase in demand for convenience food, and a reduction in leisure activity, both of which increase weight. Next, the authors considered the increase in the real price of cigarette smoking. This increase stemmed from the upward trend in money prices of cigarettes, increased restrictions on public smoking, and the success of the national anti-smoking campaign. Cigarette smokers responded to the increase in real prices by reducing demand for cigarettes, which typically leads to an increase in weight. Finally, Chou et al. considered the increased availability of convenience food. This increased supply lowers the relative price of fast

---

[32] Chou, Shin-Yi, Michael Grossman, and Henry Saffer. "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." National Bureau of Economic Research Working Paper 9247, October 2002.

food and other foods consumed outside of the home, but these foods may increase weight due to poor nutritional and high caloric value.

Chou et al. used micro-level data from the 1984-1999 Behavioral Risk Factor Surveillance System (BRFSS)[33] augmented with state level measures for per capita number of fast-food and full-service restaurants, the price of each meal in each type of restaurant, the price of food consumed at home, the price of cigarettes, clean indoor air laws, hours of work per week and hourly wage rates by age, gender, race, years of formal schooling completed, and marital status. The reduced form model used by Chou et al, using the percentage obese as the dependent variable is

$$O = O(H, F, P, S, M, EW, A, G, R) \quad (38)$$

where H is hours worked; F is family income; P is a vector of prices, including the prices of convenience foods, meals consumed at fast food and full-service restaurants, food requiring considerable preparation time, and cigarettes; S is the number of years of formal schooling completed; M is marital status; EW is the energy expended by an average person in the occupation performed by the individual; A is age; G is gender; and R includes various measures of race and ethnicity. The same independent variables are used in a second equation with BMI as the dependent variable. Three models are fitted for each of the two outcomes. The first model includes only variables measured in the BRFSS, time and time squared, the second model added the state specific variables, but excludes the time trends, and the third model includes all regressors. State specific

[33] The BRFSS consists of annual telephone surveys of persons aged 18 and older and is conducted by state health departments in collaboration with the Centers for Disease Control and Prevention. The average number of interviews per state ranged from 800 in 1984 to 3000 in 1999.

dummy variables are included in all regressions, thus the first and third models contain controls for both unobserved state and year effects.

The results from the first model demonstrate the effects of individual characteristics. Age has a U-shaped effect on BMI and obesity. BMI reaches a peak around age 57, and the probability of being obese peaks around age 45. Black and Hispanics have higher values of both variables than whites. Males have higher BMI levels than females, but females are more likely to be obese. Married and widowed respondents are heavier than single and divorced respondents. Years of schooling and household income are negatively related to BMI and obesity, thus the inverted U-shaped relationship between income and weight that is found by other analyses, is not supported by Chou et al's findings. The coefficients on time suggest that BMI and obesity are increasing at an increasing rate between 1984 and 1999.

The next model adds the state specific regressors. The coefficients for per capita number of restaurants and the real price of cigarettes are positive and significant, while the coefficients for the real prices of fast-food, food at home, and full service restaurant food are negative and significant for both BMI and the probability of being obese. Trouble arises with the restriction variables. The effects of clean indoor air laws do not reveal a consistent pattern, and the coefficients for restrictions on cigarette smoking, whether in public places or the private workplace, are insignificant. Only the effect of restrictions on elevators, public transportation, and theaters is positive and significant for obesity. The third model includes the trend terms and the state-specific variables. The full service restaurant price effect switches from negative to positive in this model, but the remaining state specific variables retain their signs, albeit reduced in magnitude. The

authors interpreted this finding as an indication of multicolinearity in the third model, and the inclusion of too many variables. In sum, the variables included in the Chou et al. model have the expected signs, and for the most part are significant, and explain a large portion of the trend in obesity. An unintended consequence of the increased opportunities for women, advances in technology, and the antismoking campaign is an increase in weight and BMI between 1984 and 1999. Their paper offers considerable insight into the effects of socioeconomic and demographic variables on obesity, along with state and time effects, without using a complex econometric structure. It serves as a starting to point when developing a model of obesity.

h. Conclusion:

In 1988, Becker and Murphy wrote the seminal paper on rational addiction. While this was not the first attempt to explain seemingly irrational behavior, including excessive drinking, gambling, and smoking, it sparked a flurry of research and hypotheses concerning such harmful decisions made by individuals. Time inconsistent agents were soon incorporated into the rational addiction framework, and economists such as Gruber and Koszegi used it to model the decision to smoke. Cutler, Glaeser and Shapiro were the first to use time inconsistency and self control issues, coupled with price reductions, to explain the growth in obesity in the United States. Several other theories followed which attempted to explain the obesity epidemic. Posner and Philipson hypothesized that technological innovations, which rendered work more sedentary, were to blame for American's expanding waistlines. Lakdawalla and Philipson empirically

tested the Posner and Philipson hypothesis, and their estimates support its predictions about the effects of technological innovations on weight. Finally, Chou, Grossman, and Saffer focus on specific societal forces that they hypothesize change the cost of nutritional and leisure time choices. In particular, their paper examined the effects of changes in relative prices on individual choices, and is a more intimate analysis of consumer responses, and the resulting increase in obesity rates witnessed in recent times.

The above papers all focus on the actions that have resulted in growth in the consumption of a harmful good. The results of these papers can be used to explain the decisions behind individual actions. Thus, they have had a profound impact on this dissertation, which attempts to explain the beliefs and decisions behind the growth rate of obesity. This is attempted first through a Bayesian model, and secondly via a behavioral model. Many authors, beginning with Viscusi, have used the Bayesian model to explain how individuals use information concerning risky behaviors. This dissertation applies this framework to the use of obesity related health risk information amongst the obese and non-obese. The behavioral model extends the examination to determine if a change in perceptions in manifested in a behavioral response.

III. Obesity Defined and Its Trends

Obesity is rapidly surpassing tobacco as the leading cause of preventable death in the United States. In December of 2001 the Surgeon General Satcher issued a call to action for Americans of all ages, races, and sexes to take notice. Over 300,000 people die each year of an obesity related disease. The CDC has long termed obesity as an epidemic, and in June 2005, they deployed a team of epidemiologists to West Virginia to study the state's debilitating obesity rate. Obesity is non-discriminatory. Table 1 displays obesity among persons age 20-74, according to sex and race. In 2000, over 30 percent of these Americans were obese. Black and Mexican females are particularly prone to becoming obese, with over 50 and 40 percent of those groups being classified as obese respectively. Tables 2 and 3, respectively, stratify the population by more refined age groups. Appendix 2 gives a graphical depiction of these trends over time. The aging population is experiencing the most rapid increase in obesity prevalence, particularly those between 60 and 69 years of age. This dissertation focuses on this age group to determine the dynamics driving their beliefs and actions.

The World Health Organization defines obesity according to body mass index, or BMI. BMI is a ratio of individual height and weight:

$$BMI = \left(\frac{weight(lbs.)}{(height(in.))^2} \right) * 703 \quad (39)$$

For example, and individual who weighs 220 pounds and is 6 feet and three inches tall has a BMI of 27.5. An individual is considered overweight if their BMI is between 25.0 and 29.9, and obese if their BMI is greater than 30.0. Appendix 3 shows BMI for various heights and weights. There are several criticisms of this metric, and its

Table 1. Changes in Age-Adjusted Prevalence of Obesity by Sex and Racial/Ethnic Group For Adults Aged 20 Years and Older*

		Obesity (BMI >=30)				
		NHANES III, 1988-1994		NHANES 1999-2000		Change, %
Sex	Racial/Ethnic Group	No.	% (SE)	No.	%(SE)	(95% CI)
Both Sexes	All**	16681	22.9 (0.68)	4115	30.5 (1.43)	7.6 (4.4 to 10.8)
Men	All**	7933	20.2 (0.72)	2043	27.5 (1.61)	7.3 (3.8 to 10.8)
	Non-Hispanic white	3285	20.3 (0.85)	946	27.3 (1.82)	7.0 (3.0 to 11.0)
	Non-Hispanic black	2112	21.1 (1.02)	374	28.1 (2.27)	7.0 (2.0 to 12.0)
	Mexican American	2250	23.9 (0.97)	538	28.9 (2.25)	5.0 (0.1 to 9.9)
Women	All**	8748	25.4 (0.95)	2072	33.4 (1.81)	8.0 (3.9 to 12.1)
	Non-Hispanic white	3755	22.9 (1.15)	885	30.1 (2.10)	7.2 (2.4 to 12.0)
	Non-Hispanic black	2490	38.2 (1.37)	420	49.7 (2.79)	11.5 (5.3 to 17.7)
	Mexican American	2128	35.3 (1.36)	567	39.7 (3.65)	4.4 (-3.4 to 12.2)

*NHANES, National Health and Nutrition Examination Survey.

**Includes racial/ethnic groups not shown separately.

***Does not meet the standard of statistical reliability and precision (relative SE > 30%)

Table 2. Changes in the Prevalence of Obesity Between NHANES III and NHANES 1999-2000 by Sex and Age*

		NHANES III, 1988-1994		NHANES 1999-2000		Change, %
Sex	Age, y*	No.	% (SE)	No.	%(SE)	(95% CI)
Both Sexes	>=20	16681	22.9 (0.68)	4115	30.5 (1.43)	7.6 (4.4 to 10.8)
Men	>=20	7933	20.2 (0.72)	2043	27.5 (1.61)	7.3 (3.8 to 10.8)
	20-29	1639	12.5 (1.14)	324	21.1 (2.42)	8.6 (3.2 to 14.0)
	30-39	1470	17.0 (1.34)	342	26.0 (2.66)	9.0 (3.0 to 15.0)
	40-49	1222	23.1 (1.57)	335	26.3 (3.28)	3.2 (-4.1 to 10.5)
	50-59	858	28.9 (1.94)	260	32.2 (4.24)	3.3 (-6.1 to 12.7)
	60-69	1179	24.8 (2.19)]	374	38.1 (3.12)	13.3 (5.6 to 21.0)
	70-79	870	20.0 (2.41)	261	28.9 (3.78)	8.9 (-0.1 to 17.9)
	>=80	695	8.0 (1.17)	147	9.6 (2.02)	1.6 (-2.4 to 5.6)
Women	>=20	8748	25.4 (0.95)	2072	33.4 (1.81)	8.0 (3.9 to 12.1)
	20-29	1669	14.6 (1.47)	304	23.3 (3.00)	8.7 (2.0 to 15.4)
	30-39	1776	25.8 (1.99)	336	32.5 (3.07)	6.7 (-0.7 to 14.1)
	40-49	1358	26.9 (1.97)	369	35.4 (3.90)	8.5 (-0.3 to 17.3)
	50-59	1005	35.6 (2.08)	284	41.2 (4.59)	5.6 (-4.5 to 15.7)
	60-69	1174	29.8 (1.53)	374	42.5 (3.46)	12.7 (5.1 to 20.3)
	70-79	985	25.0 (1.51)	250	31.9 (3.79)	6.9 (-1.3 to 15.1)
	>=80	781	15.1(1.41)	155	19.5 (3.70)	4.4 (-3.6 to 12.4)

corresponding classifications. One is that there is not distinction made between men and women, even though their body compositions are different. There is also no allowance for muscle mass. Muscle is four times heavier than fat, thus a healthy individual with a low body fat percentage may weigh enough to classify them as being obese when they are far from being unhealthy. For example Shaquille O'Neal, the towering center of the NBA's Miami Heat, weighs 325 pounds and is seven feet and one inch[34]. His BMI is 31.6, placing him in the obese category despite his high level of physical fitness. Unfortunately, the general population in not tremendously physically fit, thus the BMI measure is well suited for characterizing large, randomly selected groups, such as the HRS dataset used in this study.

Obesity is associated with a host of health complications, including high blood pressure, diabetes, heart disease, and stroke. Treating these diseases is costly. The World Bank has estimated that the cost of obesity in the United States is 12 percent of the national health care budget[35]. In 1999, the American Obesity Association commissioned a cost study by the Lewin Group[36]. They examined the costs of 15 conditions stemming from obesity, including arthritis, breast cancer, heart disease, type II diabetes, colorectal cancer, endometrial cancer, end-stage renal disease, gallbladder disease, hypertension, liver disease, lower back pain, renal cell cancer, obstructive sleep apnea, stroke, and urinary incontinence. The study indicated that the direct health care costs of obesity were $102.2 billion in 1999.

---

[34] Source: www.nba.com
[35] Source: www.worldwatch.org
[36] The Lewin Group, 1999

Obesity has continued to increase since 1999, and with it so have its costs. Considerable research has been conducted in fields ranging from physiology, psychology, and public health. Little has been done in the economics literature. Most of the existing literature in the field of economics focuses on the causes of obesity. This dissertation is the first paper to study the beliefs of the aging obese, and the behaviors subsequent to obesity related health events. More literature is needed, but this paper seeks to examine the issue from a new angle. Section IX suggests research extensions from this paper.

IV.Conceptual Model

One goal of this work is to observe how the obese and non-obese BMI groups alter their beliefs about their own mortality risks after obtaining new health information. This is the first evaluation of how perceived obesity risks change in response to personal risk information. The Bayesian risk updating framework, first proposed by Viscusi (1985), and further defined by Smith et al., is used to determine if there are differences in the responses to new information amongst BMI groups, and if these differences are attributable to the characteristics of each group.

A difference in the response to new information by BMI group may be attributed to different risk perceptions. The obese are likely to be less risk adverse because they live in the obese state despite the public health warnings about being overweight[37]. The obese may have different propensities to gather health information, and may fail to associate that information with the risky activities they undertake, due to the nature of their risk perceptions. This paper conducts tests similar to Smith et al. in order to compare the risk equivalents of the two BMI groups. The first tests if the risk perceptions of the obesity related health shocks are equal across BMI groups, and the second compares the average risk equivalents of new health information for each BMI group.

A second purpose of this study is to determine how the obese and non-obese groups alter their behavior in response to new personal health information. Risk perceptions are the basis of actions, and individual thoughts and beliefs guide utility

---

[37] Smith et al. described smokers as "risk takers" in comparison to former-smokers and never-smokers (p. 685).

maximizing decisions. The first set of estimates focuses on the beliefs of the obese, while the second examines the actions following these perceptions. This is the first behavioral response model to determine the existence and magnitude of BMI defined behavioral changes after exposure to new personal obesity related risk information. A variant of Viscusi's risk updating framework is again employed. Prior BMI level, individual characteristics, and new health information determine present BMI.

Both the risk perception model and the behavioral response model implies choice, and thus requires a utility maximizing premise. The former suggests a choice concerning the subjective longevity report, while the latter a choice concerning consumption. This study utilizes an adaptation of the Dow, Philipson, and Sala-I-Martin (DPS) (1999) model. Sam Peltzman also employed this representation in his paper "Offsetting Behavior and Medical Breakthroughs." Although the goals of these two works differ from this examination, the basic utility maximization set-up and its dynamics are applicable here. DPS focus on the sequential nature of health risks and the complementarities that arise due to risk-mitigating activities, while this paper focuses solely on the risk-mitigating activities in the presence of health risks. Many of these health risks can lead to death, but the respondent can only die once, thus, it is assumed that the individual will invest most heavily in avoiding the most significant present risk. For example, a young person may take little action to avoid congestive heart disease, and spend comparatively more on exercise. With time, given individual survival, resources will be shifted to mitigate risks more prone to the elderly, including heart failure. These expenditures might include weight loss, smoking cessation, and medical treatments.

The model couches everything in terms of individual choice between consumption and protective expenditures, and the comparative static includes a price subsidy to the protective expenditures. Instead of linking mortality risk and consumption choices as DPS do, this paper examines the evolution of the individual subjective probability of survival report, and the corresponding consumption choices. There are several implications from this framework. First, posterior survival reports evolve over time, according to the Bayesian theorem. These subjective reports may differ from the expected survival of a particular group, but new information can be utilized to minimize this difference. A second implication is that individual survival is affected by the size and composition of the consumption bundle as well as by the level of health spending. Thus, all three are chosen simultaneously.

Suppose the individual wants to maximize expected utility (EU) over a single period. In order to obtain utility, the individual must survive the entire period, and thus any utility gained or pain endured up to death is ignored. This is illustrated by the maximand:

$$EU_{it} = p_{it} \bullet \left(U_{it} | s_{it}\right) \quad (40)$$

where

EU_{it} = expected utility of individual i at time t

P_{it} = expected probability of survival of individual i at time t, given particular characteristics

$\left(U_{it} | s_{it}\right)$ = utility of individual i at time t, conditional on survival (s_{it})

The probability of survival is increased by devoting resources to health, (H), but those devoted to utility-producing consumption (X), reduce the probability. H and X are

considered composite goods. An investment in weight loss is an example of H, since it increases utility and reduces risk. Goods in H are valued mainly because they prevent illness. Driving a car is an example of X, since it increases utility and entails risk. Total resources are fixed, such that $(X + H) = \overline{K}$.

Peltzman states that the breadth of the relative tradeoff between X and H can be seen if one considers what would be included in X if survival were certain. There would be more of it, as there would be no motive to invest in medical resources or any other risk mitigating goods or services. The composition of X would also change. Peltzman suggests that food would be richer, exercise less common, and traffic restraints unnecessary. Thus, he states that H should be thought of as including a monetary equivalent of the utility forgone to enhance survival.

The problem is reduced to picking X to maximize expected utility, subject to the fixed total resources:

$$\begin{aligned} &EU_{it} = p_{it}(X,H) \bullet U(x), \\ &\textit{where}, \\ &p_X < 0, p_H > 0, U_X > 0. \end{aligned} \qquad (41)$$

The first-order condition is:

$$[(p_X + p_H)/p] + U_X/U = 0 \quad (42)$$

where the percentage gain in utility from one more unit of X balances the percentage loss in the survival probability stemming from both consuming the additional X and having fewer resources to spend on H. Alternatively, the percentage gain in survival probability due to an additional unit of weight loss and the resulting reduction in consumption of X, just balances the utility loss from one more unit of X.

An individual will invest in weight loss, if the utility attained from doing so exceeds that of the former state. For simplicity assume a simple utility function of the form:

$$U(X) = U_{it}^{BMI}$$

where U_{it}^{BMI} is the utility respondent i received from a particular BMI level in time period t. The expected utility is then:

$$EU_{it} = p_{it} * U_{it}^{BMI}$$

where p_{it} is the survival probability of individual i at time t. As an example, suppose that the following:

$$p_{i1} = 0.65$$
$$U_{i1}^{30} = 70$$

Here the survival probability associated with a BMI of 30 is 0.65 produces 70 utils of utility. Compare this to the following:

$$p_{i2} = 0.70$$
$$U_{i2}^{28} = 65$$

where the survival probability of 0.70 is associated with a BMI of 28 and yields 65 utils. This situation can be achieved over the course of a year via an additional unit of weight loss. The corresponding expected utilities are then:

$$EU_{i1} = 0.65 * 70 = 45.5$$
$$vs.$$
$$EU_{i2} = 0.70 * 65 = 48.75$$

In this instance, the individual would choose to invest in weight loss, ceteris paribus, due to the increase in the expected utility from doing so. If the expected utility from the weight loss investment were less than 45.5 in the above example, then the individual

would not invest in weight loss. Thus, the investment in weight reduction is a utility maximizing choice. This theory is used in the behavioral response model as a justification for weight loss. It may not be optimal for an individual to engage in this behavior, and it would be assumed that the choice not to lose weight to also be a utility maximizing decision.

One caveat must be made before accepting this model. From the individual's perspective, p_{it} is not observable. Thus, each period he or she must estimate his or her survival probability. The resulting subjective longevity report, $\widetilde{p}_{it}$, is observable. If $|\widetilde{p}_{it} - p_{it}| = 0$, then the individual is correctly estimating their survival, and the resulting choices of X and H are efficient. If $|\widetilde{p}_{it} - p_{it}| \neq 0$, then the decision stemming from the expected utility model is not optimal. The Bayesian updating model can be thought of as a mechanism through which the difference between the subjective report and the expected survival probability is minimized and thus $|\widetilde{p}_{it} - p_{it}| \rightarrow 0$.

The Bayesian model, as illustrated by Smith et al., suggests that the current subjective probability report is determined by several components: the prior subjective longevity report, individual characteristics, and any new information made available between the formation of the prior and the posterior longevity report. Individual characteristics and all other new information are included in the matrix $I_{it,}$, then the subjective longevity report is expressed as $\widetilde{p}_{it}(\widetilde{p}_{i,t-1}, I_{it})$. The Bayesian mechanism incorporated into this definition of the subjective probability serves as a mechanism through which $|\widetilde{p}_{it} - p_{it}| \rightarrow 0$, and the expected utility model expressed above produces the optimal choice.

V. Empirical Strategy

a. Risk-updating model (P75)

The risk updating framework, first proposed by Viscusi (1985), is used to determine if there are differences in the responses to new information amongst BMI groups, and if these differences are attributable to the characteristics of each group. This method also enables the investigator to control for other contributing factors such as household wealth, education, and gender. In this section, the response is measured via the reported subjective probability of living to age 75 or more, and new information includes socioeconomic and demographic changes, and adverse health events, both obesity and non-obesity related. The longevity report is interpreted as a measure of perceived risk.

Following Bayes, the previous wave's longevity report is termed the prior probability, which is used, along with any new information, when forming the current, or posterior longevity report. The posterior probability of individual i in time t, P_{it}, is a weighted function of a respondent's previous longevity assessment, $P_{i,t-1}$, in addition to their unobserved risk equivalent, r_{it}, implied by any new information received. Respondents are assumed to attach a level of precision to both $P_{i,t-1}$ and r_{it}, denoted as θ and γ, respectively. The precision parameters are on the [0,1] scale. The prior report is scaled by the relative information associated with it, $(\theta/(\theta+\gamma))$, while the risk equivalent, r_{it}, is weighted by the relative precision $(\gamma/(\theta+\gamma))$, and can be represented as:

$$P_{it} = \frac{\theta P_{i,t-1} + \gamma r_{it}}{\theta + \gamma} \quad (43)$$

Respondents are assumed to evaluate their likelihood of living to age 75 or more each period by using both direct and indirect signals of their current health status as personalized information[38]. The posterior report, then, encompasses all of the information available to the respondent at the time of the interview. According to this definition of P_{it}, if $\theta = 0$, the respondent ignores the prior report when forming the posterior, and relies entirely on the unobserved risk equivalent: $P_{it} = r_{it}$. Conversely, if $\gamma = 0$, the respondent's posterior report is not influenced by the unobserved risk equivalent, but entirely upon the prior report: $P_{it} = P_{i,t-1}$.

The unobservable indicator of the risk equivalent of the new information, r_{it}, is assumed to be a function of the obesity (OS_{it}) and general, or non-obesity related, (GS_{it}), health events, and changes in mobility since the previous interview, as shown in equation (44):

$$r_{it} = f(OS_{it}, GS_{it}, \Delta ADL_{it}) \quad (44)$$

It is hypothesized that these health events provide new information that the respondent uses in the Bayesian updating process to revise their longevity report. If (44) is substituted in (43), and measures of demographic and socioeconomic factors are incorporated in $z_{it1}, z_{it2}, \ldots, z_{itk}$, then we have a simple form of the model that Viscusi (1990) and Smith et al. used to describe smokers perceived risks in equation (45):

$$P_{it} = \left(\frac{\theta}{\theta+\gamma}\right) \bullet P_{i,t-1} + \left(\frac{\gamma}{\theta+\gamma}\right) \bullet f(OS_{it}, GS_{it}, \Delta ADL_{it}, z_{it1}, z_{it2}, \ldots, z_{itk}) \quad (45)$$

[38] Svenson, 1984.

There are crucial differences between Viscusi's model and that of Smith et al. Viscusi only had a cross section of data available to him, that of a survey he conducted, and thus a lag of dependent variable was not possible to include in the analysis. Additionally, the source of the updating effect could only be attributed to differences in the information available to the different demographic groups. For example, it could be said that those individuals with a higher level of education have more information about the dangers of smoking on hand than those with a lower level of education. Hypotheses such as this are troublesome, as there is no way to prove that other variables correlated with education are not driving this difference.

Viscusi's estimating equation is similar to equation (46):

$$P_t = \beta_0 + \sum_{j=1}^{k} \beta_j z_j \quad (46)$$

Thus the z_j's are assumed to capture the effect of both P_{t-1} and r_t. In comparison, Smith et al. combined all factors in r_t and z_j into one matrix, x_j, for simplicity, as can be seen in equation (47):

$$P_t = \left(\frac{\theta}{\theta+\gamma}\right) \bullet P_{t-1} + \alpha_0 + \sum_{j=1}^{k} \alpha_j x_j \quad (47)$$

Smith et al. used the first two waves of the HRS panel, thus the lag of the dependent variable was available for inclusion in the estimation equation, and the evolution of the subjective longevity report could be measured. Additionally, the role of $\left(\frac{\theta}{\theta+\gamma}\right)$ could be distinguished from the effect of new health-related information, and separate estimates of r_t and the precision weights, θ and γ were obtained.

This study takes the Smith et al. model one step further, by adding additional waves of data. The first five waves of the HRS are used to estimate a variant of (47) for the two BMI groups, obese and normal, to measure $\left(\frac{\theta}{\theta+\gamma}\right)$, α_0, and α_j for each group. This allows for the differences in the parameters between groups, and their evolution over time to be determined. Additionally, estimation biases may exist due to the lagged dependent variable in equation (47). With the additional waves of data, a fuller set of instruments is available, and may provide more reliable results. The equation used in the estimation of the Bayesian model in this dissertation is:

$$P_{it} = \left(\frac{\theta}{\theta+\gamma}\right) \bullet P_{i,t-1} + \sum_{j=1}^{k} \alpha_j x_j + \alpha_i + \varepsilon_{it} \quad , \quad (48)$$

where $\left(\frac{\theta}{\theta+\gamma}\right)$ and α_j are equivalent to those parameters in the Smith, et al. model, α_i = individual fixed effect, and ε_{it} = error term.

Two tests are conducted to understand why differences exist between the updating practices of the obese and normal sub-samples. The first tests for equality between the risk equivalents, r_{it}, of the obesity related health events across BMI groups. This helps to determine if the obese view the obesity related health events differently than the non-obese: i.e. do the obese find obesity related events to be more or less risky than the non-obese respondents do?

The estimated coefficient for the obesity related health event from equation (48), $\hat{\alpha}_{OS}$, is normalized by one minus the weight assigned to the prior longevity report for each BMI group:

$$\frac{\hat{\alpha}_{OS}^{BMI}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} \quad (49)$$

A hypothesis test for equality amongst the normalized risk measures is conducted such that:

$$\begin{aligned} H_0 &: \frac{\hat{\alpha}_{OS}^{O}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} - \frac{\hat{\alpha}_{OS}^{N}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} \leq 0 \\ H_1 &: \frac{\hat{\alpha}_{OS}^{O}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} - \frac{\hat{\alpha}_{OS}^{N}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} > 0 \end{aligned} \quad (50)$$

where O = obese risk measure, and N = normal risk measure.

The same test is conducted for the non-obesity related health events for comparison. The estimated coefficient for the non-obesity related health event from equation (48), $\hat{\alpha}_{GS}$, is normalized such that:

$$\frac{\hat{\alpha}_{GS}^{BMI}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} \quad (51)$$

The hypothesis test becomes:

$$\begin{aligned} H_0 &: \frac{\hat{\alpha}_{GS}^{O}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} - \frac{\hat{\alpha}_{GS}^{N}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} \leq 0 \\ H_1 &: \frac{\hat{\alpha}_{GS}^{O}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} - \frac{\hat{\alpha}_{GS}^{N}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} > 0 \end{aligned} \quad (52)$$

Rejection of the null hypothesis suggests a difference between the normalized risk measures by BMI group.

A second test uses a reduced form expression for the risk equivalent of the new information that is assumed communicated through the adverse health events. This risk equivalent is calculated at the mean, such that the reduced form expression becomes:

$$\bar{r} = \frac{\alpha_0 + \sum_j \hat{\alpha}_j \bar{x}_j}{1 - (\hat{\theta} / \hat{\theta} + \hat{\gamma})} \quad (53)$$

where the denominator is one minus the coefficient for the prior longevity perception, and the numerator includes information from all other sources, including the adverse health events. The estimated updating model, captured in the estimated coefficients, is applied to the mean values of the BMI group's observed experience to estimate the group mean risk equivalent, $\bar{r}$. These estimates allow the effects of each group's weights for information to be separated from their actual health-related experiences, and for comparisons to be made. For example, using this technique, it can be shown, on average, how the obese would interpret new health information, given the health experiences and risk updating methods of the normal sample.

These results yield further insight into how the obese interpret information. This test makes it possible to determine if the obese interpret personal experience differently than the normal sample interprets the same average experience. If a difference is found, it can be said that the obese do use and view information differently then the normal group, giving a reason for the initial findings that the obese respond differently to adverse health events than the non-obese. That reason being, there are differences in the risk perceptions for the same events amongst the BMI classifications.

b. Behavioral model (BMI)

The second model, outlined here, takes the risk-updating model above a step farther. Instead of using the Bayesian updating model to determine how individuals use new

information to form their current risk perception, the model is incorporated into a behavioral response model. The behavioral response model is used to determine if respondents use new information about risk to alter their behavior, the direction of the change, and does the response differ across BMI groups. The behavioral response measure used in this analysis is BMI. Height is not assumed to change over the eight years of the survey, and a change in BMI is interpreted as a change in weight.

The behavioral response model is:

$$BMI_{it} = \left(\frac{\theta}{\theta+\gamma}\right) \bullet BMI_{i,t-1} + \sum_{j=1}^{k} \alpha_j x_j + \alpha_i + \varepsilon_{it} \quad (54)$$

Again, the prior report, in this case, prior BMI report, is scaled by the relative information associated with it ($\theta/(\theta+\gamma)$), while the risk equivalent, r_{it}, is again scaled by the relative weight, $\left(\frac{\gamma}{\theta+\gamma}\right)$. Care should be taken when interpreting $\left(\frac{\theta}{\theta+\gamma}\right)$, as the response measure is no longer a subjective probability report as in the first model, but an objective measure. θ is the influence of the respondent's prior BMI report on the current BMI report. A positive coefficient suggests that there is a positive relationship between the reports, or that BMI levels increase over time. In contrast, a negative coefficient indicates that there is negative reinforcement, and that the respondent is losing weight over time. As in equation (48), α_i is an unobserved fixed effect, and ε_{it} is the error term.

VI. Data

a. HRS

This study uses the RAND HRS Version B dataset, which is a "cleaned, processed, and streamlined"[39] version of the Health and Retirement Survey (HRS). Funding for the HRS was provided by the National Institute on Aging at NIH, the Social Security Administration, the Department of Labor Pension and Welfare Benefits Administration, the Office of the Assistant Secretary for Planning and Evaluation at DHHS, the State of Florida Department of Elder Affairs, the NIH Office of Research on Minority Health, and the NIH Office of Research on Women's Health. The HRS is a national panel survey of individuals, aged 51 to 61 in the baseline survey year, 1992, and their spouses. The goal of the survey is to provide data to promote research and analysis of retirement, health insurance and savings decisions of the older US population. The dataset includes information regarding demographics, income, assets, health, cognition, family structure, health care utilization and costs, job status, and savings and longevity expectations. The survey is conducted every two years, with the most recent publicly available wave being from 2000, making 5 waves available at the time of this paper. In 1992, the population included 12,562 respondents from 7,702 households, with an over-sampling of Hispanics, Blacks, and residents of Florida. Weights are applied to make it a representative population. The baseline surveys were conducted face to face, and follow up interviews were conducted over the phone.

---

[39] StClair, Patricia, et al. "RAND HRS Data Documentation," Version B, April 2003, prepared for the Social Security Administration, pg. 2.

The RAND Center for the Study of Aging took the "raw" HRS data and created the RAND HRS Data files, which includes the first four waves of the HRS final release files and the preliminary release file of the fifth wave. Several other cohorts are added to the HRS group in the RAND Data file, including Children of the Depression Age (CODA), War Babies (WB), and AHEAD. This dataset is easier to use than the original "raw" HRS dataset. As a consequence of being easier to use, the RAND version has less detailed information: a sacrifice of computational ease. The necessary variables are still included to conduct this analysis, and thus the most recent version of construct, Version B, is used.[40]

The HRS is a suitable data source for this study. Obesity is a non-discriminatory epidemic, spanning all age groups, races, and genders. Tables 2 and 3 display NHANES[41] data across several age cohorts, including the HRS age set, from 1988 to 2000. According to Table 2, the age groups 50-59 and 60-69 experienced the greatest obesity prevalence and growth rates for both men and women. Men experienced an increase of 13.3 percent, while women witnessed a 12.7 percent increase, compared to 3.2 and 8.5 percent for men and women aged 40-49, respectively. Over the five waves of the survey included in this study's analysis, individuals advanced from ages 51-61 to 59-69, placing the respondents within these troubled aging cohorts. This finding suggests that the age cohort surveyed in the HRS is an especially interesting one for the study of

---

[40] Since the onset of this analysis, more updates have been made, such as RAND HRS Version C, but the alterations do not affect the information used here, so that Version B is sufficient.

[41] National Health and Nutritional Examination Survey (NHANES) provides information concerning the health and nutritional status of the civilian, non-institutionalized US population, residing in the 50 States and the District of Columbia. Each sample is nationally representative, but in 1999 the NHANES design shifted from a periodic to a continuous collection method. Participants provided information through several methods, including a physical exam at a mobile exam center and telephone interviews. More information concerning NHANES can be found at www.cdc.gov/nchs/nhanes

obesity. The aging population is the group with the fastest growing obesity prevalence in the United States. Obesity is a condition of increasing incidence for the entire population, but it is particularly pervasive amongst the aging population. No economic study to date has focused on this age cohort.

Appendix 1 displays trends in the prevalence of obesity for adults aged 20-74 from 1960 to 2000, a longer time period than that shown in Table 2. Clear distinctions appear between men and women over the 40-year period. Between 1960 and 1988, women between the ages of 60 and 74 years of age experienced the greatest obesity prevalence, although it declined throughout that period. By the 1976-1980 period, all women displayed a remarkable increase in obesity prevalence, led by the 40-59 year old age group. This growth increased through 2000, although the 60-74 age cohort were the forerunners. Men aged 40-59 displayed the greatest obesity prevalence from 1960 to 1999, at which time their older peers, the 60-74 year olds, surpassed them. Obesity prevalence increased throughout all time periods measured, although its prevalence was less amongst men than women.

Adverse health events are the primary source of new information included in the Bayesian updating process. The RAND-HRS dataset's indicators of these events lack considerable, and possibly important details. As is, the included adverse health events are:

- Heart disease
- Lung disease
- Diabetes
- Stroke

- High Blood Pressure
- Arthritis
- Cancer

For many of these complications, especially heart disease and cancer, there are different forms of the disease, some with different sources, causes, and level of severity associated with them. For example, an individual may adjust their assessment of living to age 75 differently when diagnosed with a heart attack versus heart angina. The RAND-HRS dataset incorporates both of these complications in the heart disease measure. The raw HRS dataset contains considerably more detail than the RAND version.

Details of the health events are infused into the RAND dataset from the raw HRS dataset to illicit more accurate measures of the updating process resulting from the exposure to new information. The alterations include subdividing heart disease into heart angina, heart attack, and congestive heart failure; diabetes into diabetes treated with insulin and diabetes treated with oral medication; and stroke into stroke requiring medication and stroke resulting in permanent difficulties and problems. The list of adverse health events then becomes:

- Heart angina
- Heart attack
- Congestive heart failure
- Lung disease
- Diabetes- insulin
- Diabetes-oral medication
- Stroke-medication

- Stroke-problems
- High Blood Pressure
- Arthritis
- Cancer[42]

A validity filter must be applied before reports of the adverse health events can be considered reliable sources of new information[43]. The individual must either be on medication for the complication, seeing a physician for it, or both to be considered as having had an adverse health event in the dataset. The following filter applied for angina serves as an example. Angina is included in the dataset as a valid health event if the respondent gives a positive response to all of the following questions:

1. Has a doctor ever told you that you had a heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems?
2. Do you currently have angina or chest pains due to your heart?
3. Are you taking or carrying any medications because of your chest pains?
4. During the last 12 months, have you seen a doctor for any of your heart problems?

Filters such as the one used for heart angina elicit confidence that the health event did actually occur, and that the respondent is not reporting a self-diagnosis. Details of the remaining filters are shown in Appendix 4.

Several sorting criteria are used to clean up the RAND-HRS dataset prior estimation. The goal is to simultaneously eliminate erroneous entries without rendering

[42] Details of the different types of cancer diagnoses render sample sizes that are too small to be used with any degree of confidence.
[43] Appendix 4, Table 5 show these filters.

the population non-random, and to construct the most informative dataset for the Bayesian updating and behavioral models. The focus wave is the baseline wave, as all behavioral information builds upon this year. Therefore, all respondents who are not alive in the initial wave are eliminated. Past illnesses increase the probability of future adverse health events, and may result in an updating of the subjective longevity perceptions before the individual is exposed to the included prior new information. Respondents who do not enter the study with a pre-existing health condition are eliminated. The resulting sample is healthy in the baseline wave, and any new health information yields a fresh, unbiased response[44].

Finally, the remaining respondents are divided according to their body mass index (BMI)[45] classifications. The World Health Organization has suggested the following categories[46]:

Underweight: BMI < 18.5

Normal: 18.5<= BMI< 25

Overweight: 25<= BMI < 30

Obese: 30<= BMI < 40

Severely Obese: BMI >= 40

Two categories are of interest for this study: the normal and obese groups. The overweight, obese, and severely obese are condensed into an obese group in this

---

[44] Respondents with missing observations are eliminated from the panel.
[45] See Appendix 3 for a BMI table.
[46] World Health Organization. Obesity: Preventing and Managing the Global Epidemic. Geneva: World Heath Organization, 1997.

dissertation. Including the overweight and the severely obese with the categorically obese respondents both increases the sample size and reduces the arbitrary nature of the BMI classification cut-off points. The obese are compared to the normal group in the estimation process. The HRS asks respondents to report their weight in each wave, and their height in the first two waves. A possible bias that may result from this classification system is that self-reported weight is likely to be an underestimate of actual weight, especially at the upper tail of the BMI distribution. This has the potential to lead to inaccuracies in estimation. Clinical measures of weight are preferable, but if this study finds behavioral differences between the groups, and the obese are likely to underreport their weight to a greater degree than the normal, it can be inferred that the estimates also underreport the behavioral differences. The possible bias should be kept in mind when interpreting any results based upon this dataset. The response attrition from these filters and sorting criteria results in 1,319 observations for the obese sample and 588 for the normal sample, and is detailed in Appendix 5.

46 percent of the healthy HRS sample is obese over the years 1991 to 2000. NHANES comparable time periods, 1988-1994 and 1999-2000 show that 22.9 percent and 30.5% of those samples are obese, respectively. A greater share of the obese HRS sample is male, compared to the dominance of the women in the NHANES study. Additionally, whites hold a greater share of the obese sample than that of the NHANES. The filters applied to the HRS dataset in this study are a likely cause of some of the differences between the HRS and NHANES samples. Additionally, the HRS focuses on individuals aged 51-61, and follows the same respondents over time. The NHANES, in contrast, interviews qualified individuals between the ages 20 and 74. When the

NHANES is paired down to more refined age groups, as in Tables 2 and 3, the prevalence rates begin to converge.

b. Subjective Probability of Living to Age 75

The subjective probability of living to age 75 or older is the agent through which the updating process operates in the Bayesian model. The measure is constructed using the respondents' answers to the following question:

> *"Using any number from zero to ten, where 0 equals absolutely no chance, and 10 equals absolutely certain, what do you think are the chances that you will live to be 75 or more?"*

The respondents are presented a scale similar to Table 9 below as an aide for answering the subjective longevity question.

Table 3: Subjective Probability Scale

00	01	02	03	04	05	06	07	08	09	10

Responses are rescaled to the [0,1] interval to represent a probability[47]. In the dataset used in this study, the normal sample reports an average subjective probability of 0.68,

[47] Smith et al. used this technique in their analysis.

compared to 0.66 for the obese sample, a difference of 4 percent.[48] This trend continues when the data are further stratified by gender. Normal men are one percentage point more optimistic than their obese counterparts (67.1 versus 66.1 percent). An even wider gap exists between the female BMI groups. Normal women report that they are 71.3 percent sure that they will survive at least until age 75. Only 66.7 percent of the obese women believe this to hold for them. It is expected that women would be more optimistic about their longevity then men, as it is widely known that women live longer than men. The normal group exhibits the expected difference in the longevity report by gender, but the obese gender reports are not significantly different.

According to the Behavioral Risk Factor Surveillance Study (BRFSS), the life expectancy for all individuals at age 65 in 2000 is 82.9 years of age.[49] By gender these ages are 81.3 for men and 84.2 for women. Thus it is expected that on average, all men and women will live to be age 75 or older. Objective measures such as those included in the BRFSS exclude personal information that may influence the subjective probability report. Some of this information is observable to the econometrician, including lifestyle factors, while other aspects are not, including genetic traits. Thus, when studying the risk updating process of individuals, a subjective measure, such as that included in the HRS, is preferable to an objective measure, including the BRFSS.

There has been substantial controversy concerning the validity of the HRS subjective longevity perception measure. Smith, Taylor and Sloan[50] tested whether

---

[48] See Appendix 8, Table 8 for summary statistics.
[49] Behavioral Risk Factor Surveillance System, www.cdc.gov/brfss
[50] Smith, V. Kerry, Donald H. Taylor, and Frank A. Sloan. (2001). "Longevity Expectations and Death: Can People Predict their Own Demise?" *American Economic Review*, 91(4), 1126-1134.

longevity expectations predict mortality at the individual level using the first four waves of the HRS dataset. They found evidence suggesting that the HRS subjective probability measure accurately predicts future mortality among the same individuals, and that observed deaths are "signaled" through lower longevity expectations respondents report in earlier interviews. The authors also consider the HRS variable to be valid because respondents update their subjective probabilities of living to 75 in plausible directions following health shocks.

Picone, Sloan and Taylor[51] state that a subjective estimate of longevity is superior to an objective longevity estimate, such as life table estimates for two reasons. First, the individual knows much more about his/her own health than is incorporated in the life tables. Second, the individual bases decisions on such beliefs. Thus, the subjective longevity perceptions capture characteristics unobservable to the econometrician, and are superior for estimation purposes such as this paper, which seek to understand the beliefs behind individual actions. Smith, Taylor and Sloan's work supports this idea that individuals know more about their own health, and that self-reported measures are superior to objective measures.

Hurd and McGarry[52] test the validity of the subjective longevity probability in the HRS by looking at its covariates over time. Although these authors only conduct these tests on the first two waves of the dataset, the tests were repeated for all five waves by

---

[51] Picone, Gabriel, Frank Sloan, and Donald Taylor. (2004). "Effects of Risk and Time Preference and Expected Longevity on Demand for Medical Tests." *Journal of Risk and Uncertainty* 28(1), 39-53.
[52] Hurd, Michael D., Kathleen McGarry. (1995). "Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study," *Journal of Human Resources* 30, Special Issue on the Health and Retirement Study: Data Quality and Early Results, S268-S292; Hurd, Michael D., Kathleen McGarry. (2002). "Predictive Validity of Subjective Probabilities of Survival," *Economic Journal* 112, 966-985.

this author, which support Hurd and McGarry's findings. The subjective longevity probabilities covary as expected with education, income, self-report of health, drinking and smoking status, and parental longevity in all five waves, again suggesting that the HRS measure behaves as a valid measure of expected survival.

Several authors criticize the use of subjective probabilities, including those found in the HRS. Lichtenstein et al. (1978) were the first to raise questions about the validity of such subjective evaluations[53]. Viscusi and Hakes'[54] (2002) empirical analysis questions whether the HRS captures a true subjective probability. They are concerned that the HRS probabilities may not be a consistent representation of the true probability. Picone et al. address this issue. They state that this is only a concern if the subjective probabilities are compared to objective probabilities. This dissertation attempts to compare the subjective probabilities of the obese and the non-obese, and their evolution over time. No comparison is made with objective measures, and thus the literature supports using the HRS subjective measure a Bayesian updating framework

[53] Lichtenstein, Sarah, Paul Slovic, Baruch Fishoff, Mark Lyman, and Barbara Combs. (1978). "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory* 4, 551-579.
[54] Viscusi, W. Kip and John K. Hakes. (2003). "Risk Ratings that do not Measure Probabilities," *Journal of Risk and Insurance* 6(1), 23-43.

c. Health Events[55]

1. Angina Pectoris[56]:

Angina[57] refers to pain caused by restricted blood flow in the arteries that supply the heart. Angina typically initiates during exercise, and is especially likely to occur when walking after a meal, or in cold, windy weather. Anger or stress can exacerbate the condition. Angina feels like a tightening across the upper chest. The pain can radiate to the neck, throat, or arms, and lasts for several minutes. In addition to the pain, breathlessness, sweatiness, and a sense of fear are common symptoms.

Most angina is due to disease of the coronary arteries (atherosclerosis), a consequence of the build-up of fatty deposits or plaque in the arteries. The arteries become narrower due to the build-up, and the heart muscle is no longer able to receive enough blood carrying oxygen and nutrients especially when extra demands are made on it through exertion. The narrowing of the aortic valve, anemia, fast and abnormal heart rhythms, and diseases of the heart muscle can also cause angina.

Angina occurs most often in older adults. Individuals more prone to angina include those who:

- Are smokers
- Have high cholesterol
- Have high blood pressure
- Diabetic

---

[55] See Appendix 9 and 10, Tables 9 and 10 respectively, for a summary of the risk factors of each health events and their treatments.
[56] Source: American Heart Association's webpage, www.americanheart.org
[57] Angina comes from the Latin word *angerer*, which means to strangle.

- Are inactive
- Have a family history of angina

It is estimated that 6,400,000 Americans suffer from angina, and that there are 400,000 new cases of stable angina each year[58]. The estimated age-adjusted prevalence of angina in women age 20 and older is 3.5 percent for non-Hispanic white women, 4.7 percent for non-Hispanic black women and 2.2 percent for Mexican-American women. The prevalence percentages for their male cohorts are 4.5, 3.1, and 2.4 respectively.

The HRS dataset reports incidence rates close to those at the national level discussed above. 2 percent of the normal HRS sample reports diagnosis of angina, compared to 3.5 percent of the obese group. Thus, angina is 75 percent more prevalent amongst the obese than the normal sample, further suggesting that it is an obesity related event. Differences also exist across genders. 3.6 percent of obese men suffer from angina in the HRS dataset, compared to 2.5 percent of their normal counterparts. Women report a lower incidence of the condition than their male BMI peers, as only 1.2 percent of normal women and 2.9 percent of obese women report suffering from angina.

Typically, conditions that make angina more likely to occur are treated in addition to the angina itself. Thus, physicians will recommend treatments for high cholesterol and blood pressure, and diabetes, if present, and suggest lifestyle changes, including smoking cessation and increased physical activity. Angina is classified as an obesity-related health event since the prevalence of the complication increases with weight, and because weight loss is considered a treatment.

---

[58] Framingham Heart Study, National Heart, Lung, and Blood Institute.

2. *Heart Attack:*

Coronary heart disease also causes heart attack. While angina describes the pain caused by restricted blood flow in the arteries that supply the heart, heart attack, or myocardial infarction, occurs when the blood supply to part of the heart muscle (myocardium) is slowed or stopped. This can occur when one or more of the coronary arteries supplying the heart muscle is blocked, usually due to artherosclerosis. The plaque can rupture or tear, creating a blood clot that can block the artery. This blockage can result in a heart attack. If the heart's blood supply is cut off for more than a few minutes, muscle cells suffer permanent damage and can die. This can kill or disable someone, depending upon the extent of the damage. Angina and myocardial infarction (MI) are both caused by atherosclerosis. Thus the risk factors for angina and heart attack are the same. A heart attack is an acute health event that is classified in this analysis as an obesity-related condition.

In 2002, 494,382 deaths in the United States, one of every 5 deaths, were attributed to an MI. There are 1,200,000 new and recurrent cases of MI per year, and 41 percent of the people who experience a MI in a given year die from it[59]. Within the HRS dataset, 2.0 percent of the obese sample reports a diagnosis of heart attack over the last two years, while only 1.3 percent of the normal group does, making heart attack 54 percent more prevalent amongst the HRS obese. This gives further support for classifying heart attack as an obesity-related condition amongst this sample. Obese men

---

[59] National Heart, Lung, and Blood Institute's Atherosclerotic Risk in Communities (ARIC) Study, 1987-2000.

report the highest incidence of heart attack when the data is sorted across BMI and gender groups. Within the normal group, there is a 1.3 and 1.2 occurrence for men and women respectively, compared to 2.2 and 1.2 percent for the obese men and women.

3. *Congestive Heart Failure (CHF)*[60]*:*

"Heart failure" means that the heart is not pumping as efficiently as it should to deliver oxygen-rich blood to the body. CHF occurs when the heart's weakened pumping action results in a buildup of fluid called congestion in your lungs and other body parts. The condition develops slowly, and may go undetected for years. The heart attempts to compensate for its weakened state by enlarging, and forcing itself to pump faster to move more blood through the body.

CHF symptoms depend upon which side of the heart is affected. If the left side of the heart is not working properly, blood and fluid back up in the lungs, and the individual may feel short of breath, tired, and develop a cough. Right-sided failure results in a buildup of fluid in the veins due to the slowed blood flow. The feet and ankles begin to swell, and condition called edema. Edema can spread to the lungs, liver, and stomach, and can lead to kidney failure. Additional symptoms are:

- Shortness of breath especially while lying flat
- Fatigue, weakness and inability to exercise or perform other physical activities
- Weight gain from fluid retention

---

[60] Source: American Heart Association's web-page www.americanheart.org

- Chest pain
- Loss of appetite
- Swelling of neck veins
- Skin feeling cold and sweaty
- Rapid and irregular pulse
- Restlessness, confusion, and loss of attention span and memory

Therapies for CHF focus on easing the workload of the heart. Treatments include lifestyle changes, medication, transcatheter interventions, and surgery. The recommended lifestyle changes are[61]:

- Smoking cessation
- Control of HBP, cholesterol levels, and diabetes
- Limiting intake of fluids, saturated fats and salt
- Increase amount of aerobic exercise

Medication has also proven to improve heart function and the ability to perform physical activity. These medications include diuretics, beta-blockers, ACE inhibitors, and calcium channel blockers. Percutaneous coronary interventions such as angioplasty and stenting improve outcomes for CHF patients. Finally, congestive heart failure may progress to a stage that a surgical procedure is recommended. These include pacemaker insertion, coronary artery bypass surgery, and heart transplantation.

Close to five million Americans are living with congestive heart failure and between 400,000 and 700,000 new cases are diagnosed each year. About 250,000

---

[61] Sleep apnea is a recently identified cause of CHF. No data was available for this problem. Obesity is a known cause of sleep apnea, and a large share of CHF in the obese may be due to this condition.

million individuals die annually from the disease[62]. 0.34 percent of the HRS normal group reports a diagnosis of congestive heart failure in the past two years, compared to 0.94 percent of the obese. Thus, CHF is 57 percent more common amongst the HRS obese, yielding support for classifying it as an obesity-related complication.

4. *Diabetes*[63]*:*

Diabetes mellitus killed 73,249 Americans in 2002, 46.8 percent male and 53.2 percent female. It contributes to almost 200,000 deaths in the U.S. each year. Type II diabetes is the most common form of the disease, and is often linked with obesity and can be controlled with diet and exercise. In a mild form, the disease can go undetected for years, and lead to additional chronic illness, including cardiovascular disease. Type II diabetes is a progressive disease that develops when the body does not efficiently use insulin produced, or insulin resistance.

High-risk groups for Type II diabetes are:

- Over 45 years of age
- Overweight
- African-American, Latino/Hispanic American, Asian American, or Pacific Islander
- A family history of the disease

---

[62] Source: Mayo Clinic, web-page www.mayoclinic.com
[63] Source: American Heart Association, web-page www.americanheart.org

Patients with diabetes must control their blood glucose levels daily to prevent or delay complications such as cardiovascular disease. This can be accomplished via oral medications such as thiazolidinediones, meglitinides, and biguanides. Such oral medications are not useful for a Type I diabetic, whose pancreas is no longer capable of producing insulin. Insulin may also be needed for hard to control Type II diabetes. In addition to medication, diet and exercise are crucial to controlling diabetes.

Cardiovascular disease is the most common, and one of the most serious, complications of diabetes, which increases risk of heart attack and stroke. 65 percent of people with diabetes die of one of these two conditions[64]. Diabetes damages the blood vessels, including those that supply blood to the brain and heart. Atherosclerosis can reduce or even stop the blood supply and drive up blood pressure. Having diabetes makes an individual two to four times as likely to have a stroke or heart attack than someone without diabetes. Curiously, close to 70 percent of people with diabetes are not aware of these increased risks. Many diabetics believe that foot or leg amputation and blindness is their biggest threat[65].

1.7 percent of the HRS obese group reports a diagnosis of diabetes for which oral medication is needed in the past two years, while only 0.08 percent report a diagnosis of diabetes for which insulin is needed. It can be concluded that at least 1.7 percent of the obese sample is diagnosed with Type II diabetes. Only 0.51 and 0.09 percent of the normal group reports a diagnosis of diabetes requiring oral medication and insulin respectively. It is not surprising that 234 percent more obese respondents are

---

[64] Source: The Mayo Clinic's web-page www.mayoclinic.com
[65] Ibid.

diagnosed with orally medicated diabetes than the normal group, since only Type II diabetes can be treated with oral medication, and Type II diabetes is known to be more prevalent amongst the overweight and obese. Type I diabetes is not linked to obesity, and thus there is not a statistically significant difference between the diagnosis reports across BMI groups. Diabetes is categorized as an obesity-related health event in this analysis[66].

5. *Stroke:*

A stroke is an acute condition, caused by a rupture or blockage of an artery in the central nervous system. In middle aged- or older women, close to 70% of strokes are thromboembolic, or caused by a blockage from a blood clot, 15% consist of intracerebral hemorrhage, and 10% of subarachnoid hemorrhage[67]. The portion of the brain that is deprived of oxygen depends upon where the ruptured or blocked artery leads. The lack of oxygen can result in permanent brain damage, disability, and death.

Risk factors for stroke include:

- High blood pressure
- Tobacco use and excessive alcohol consumption
- Diabetes mellitus

[66] A closely related health complication caused in part by obesity is the metabolic syndrome. It is characterized by a group of metabolic risk factors in one person: central obesity (excessive fat tissue in and around the abdomen), atherogenic dyslipidema, raised blood pressure, insulin resistance or glucose intolerance, prothrombotic state, and proinflammatory state. The underlying causes of this syndrome are overweight/obesity, physical inactivity, and genetic factors. These individuals are at an increased risk of coronary heart disease, other diseases related to plaque buildups in artery walls, and type 2 diabetes. More information can be found at www.americanstroke.org

[67] Source: American Stroke Society, www.americanstroke.org

- Heart disease
- High blood cholesterol
- Physical inactivity and obesity
- Age, sex, gender, and race
- Prior stroke or heart attack

Stroke prevention includes:

- Control of high blood pressure and diabetes
- Treating established heart disease
- Low to moderate alcohol consumption
- Smoking cessation
- Weight loss

Further research is needed to determine the relationship between blood cholesterol levels and prevalence of stroke, and the excess risk of stroke among black women.

Surgery, medications, hospital care, and rehabilitation are all used as treatment for stroke. Cerebral angioplasty is a technique in which balloons, stents, and coils are used to widen the brain's blood vessels. Plasminagen activators (tPA) are used to treat strokes caused by blood clots by dissolving the clot and restoring blood flow to the brain. For maximum benefit, the drug should be taken within three hours of the onset of stroke symptoms.

In the United States, stroke is the third leading cause of death and is the leading cause of adult disability[68]. Each year about 700,000 American suffer a stroke, and 160,00 of them die. The number of stroke related deaths has been falling over the past 20 to 30 years. This is largely due to the improvement in the control of certain risk factors, including smoking, high blood pressure, and high cholesterol. The increase in the incidence of obesity could stagnate these improvements, as obesity is also a risk factor for stroke.

0.5 percent of the obese sample reports a diagnosis of a stroke in the last two years, compared to 0.6 percent amongst the normal. 0.5 percent of the obese describe problems resulting from the stroke, while only 0.3 percent of the normal report these complications. It appears that, although there is not a statistically significant difference in the incidence of stroke between the BMI groups, the obese have more complications stemming from the stroke. Because of the medical evidence, and the greater number of post-event complications, stroke is classified as an obesity-related health event in this study.

6. Lung Disease[69]:

Lung disease typically refers to the pair of complications, chronic bronchitis and emphysema, collectively known as Chronic Obstructive Pulmonary Disease, or COPD. Chronic bronchitis is the inflammation of the lining of the bronchial tubes, which restricts

---

[68] Source: www.mayoclinic.com
[69] Definitions obtained from the American Lung Association website, www.lungusa.org.

airflow to and from the lungs and causes heavy mucus to be coughed up. Bronchitis can be acute or chronic. Chronic bronchitis by definition is defined as the presence of a mucus producing cough most days of the month of a year, for two successive years, without any other explanations for the cough. Cigarette smoking is the most common cause of chronic bronchitis, but it can also stem from exposure to dusts and fumes in the workplace, and other forms of air pollution. Chronic bronchitis may precede or accompany pulmonary emphysema. Emphysema is a condition in which the walls between the alveoli within the lungs weaken and break, after the loss of their elasticity. Air is trapped in the air sacs, and the ability of the lungs to exchange oxygen and carbon dioxide is impaired.

COPD is the fourth leading cause of death in the United States, and could surpass stroke by 2020[70]. An estimated 16 million people are diagnosed with COPD in the United States, and perhaps as many as 14 million are left undiagnosed. People over the age of 50 are more likely to be considered disabled due to COPD, and about 1.5 million emergency room visits were due to the disease. Smokers are 10 to 15 times more likely to be disabled by emphysema than non-smokers.

0.5 percent of the HRS normal sample reports a diagnosis of lung disease in the last two years, relative to 0.7 percent amongst the obese. This is a 40 percent greater incidence amongst the obese, although the difference is not statistically significant. Lung disease is classified as a general health shock in this study both due to the lack of medical evidence, and the absence of support from the HRS sample categorizing it as an obesity-related condition.

---

[70] Source: www.copd-international.com

7. <u>High Blood Pressure</u>:

Blood pressure is the force in the arteries when the heart beats and when the heart is at rest. These two types of pressure are known as the systolic and diastolic pressures, respectively. Blood pressure is reported as systolic/diastolic pressure. The recommended blood pressure level is 120/80 or less. High blood pressure is defined as a blood pressure greater than or equal to 140/90[71] for adults, and lower for people with other conditions, such as diabetes and kidney disease. Risk factors for high blood pressure are:

- Being overweight
- Physical inactivity
- Excessive consumption of alcohol
- Smoking
- Unhealthy diet
- Stress

Being black, over the age of 55 and having a family history also increase the likelihood of high blood pressure. Some medications, including some antidepressants, cold medications, and some oral contraceptives can raise your blood pressure, as can some medical conditions such as kidney disease, Cushing's disease, and sleep apnea. A healthy lifestyle is the ideal way to prevent and control high blood pressure. It may be

[71] Recent recommendations have lowered this substantially.

necessary to add blood pressure medications, including diuretics, Beta-blockers, Alpha-blockers, and vasodilators.

High blood pressure is a common condition.[72] About 65 million Americans aged 20 and older have high blood pressure. Of those with high blood pressure, 30 percent do not know that they have it. Of those who know they have the condition, 11 percent are not on therapy, 25 percent are on inadequate therapy, and 34 percent are on adequate therapy. The cause of 90-95 percent of the cases is not known, although it is easily detected and usually controllable. From 1992 to 2002 the death rate from high blood pressure increased 26.8 percent, and the number of deaths by 56.6 percent. In the HRS dataset, 3.8 percent of the normal population reports a diagnosis of high blood pressure in the last two years, whereas 4.1 percent of the obese make such a claim, a difference of only 8 percent. High blood pressure is cataloged as obesity related in this study because medical evidence links the two, and because weight loss can reduce the patients' blood pressure.

8. Arthritis[73]:

The term arthritis refers to a group of more than 100 medical conditions that effect almost 70 million adults and 300,000 children in the United States[74]. The most common form of arthritis is osteoathritis (OA). It is most prevalent in people over sixty years of age and the overweight. All forms of arthritis affect the musculoskeletal system,

---

[72] Source: American Heart Association, web-page www.americanheart.org
[73] American Heart Association, www.americanheart.org
[74] Statistic and description from the Arthritis Foundation's webpage, www.arthritis.org

primarily the joints. Arthritis-related problems include pain, stiffness, inflammation and damage to the joint cartilage and the surrounding tissue. This damage can result in joint weakness and instability that can interfere with activities such as walking and climbing stairs. Arthritis is considered an obesity related disease in this study, since most cases of arthritis are diagnosed as osteoarthritis. Osteoarthitis is a weight bearing condition. Being overweight and obese places excessive weight on the joints of the knees, hips, feet and spine, and can exacerbate in osteoarthritis.

Arthritis treatment focuses on:

- Relief of pain
- Minimizing joint stiffness
- Reduction of inflammation
- Preserving muscle and joint function
- Maintaining a normal lifestyle
- Minimizing side effects of arthritis medication

The treatment methods do not offer a cure for arthritis, but help patients maintain their quality of life, and include rest, anti-inflammatory medications, physical therapy, and drug therapy to slow the progression of the disease.

2 percent of the obese report a diagnosis of arthritis over the past two years in the HRS survey data, versus only 1 percent of the normal sample, a difference of 100 percent. Nationwide, close to 70 million people have some form of arthritis or chronic joint problems. Rheumatic diseases are the leading cause of disability amongst the elderly population. The CDC reports that people who are overweight or obese have consistently more arthritis diagnoses than the non-obese. In fact, 66 percent of adults

diagnosed with arthritis were overweight or obese in 2002[75]. Weight loss of as little as eleven pounds can reduce the risk of developing osteoarthritis of the knees by 50 percent[76].

d. Summary Statistics

Table 4 displays the summary statistics for the obese and normal samples. Large portions of these statistics have been discussed in previous sections. This section examines the remaining variables and their trends. The dependent variable for the Bayesian model is the subjective probability of living to age 75 or more, while it is BMI in the behavioral model. The means for these variables and the explanatory variables are shown across all waves for a general analysis. The summary statistics are generated after all sorting criteria are carried out, and the data set is ready for estimation.

The summary statistics across all waves reveal both remarkable similarities and differences between BMI groups. Marital status is almost identical across groups with an astounding 99 percent reporting that they are married. This suggests that respondents reporting no adverse health conditions in the initial wave are almost certain to be married. The obese and normal samples are also comparable according to their accumulated assets and the number of drinks consumed per day. The normal sample has attained a higher level of education, and is more likely to be white and unemployed than the obese group. Males, and minority groups more heavily dominate the obese group than their normal

---

[75] Unpublished CDC data (Source: 2002 NHIS, from the CDC website www.cdc.gov)
[76] *Arthritis Rheum* 1998; 41(8): 1343-1355 (Data Source: Framingham Osteoarthritis Study)

peers. The obese also exercise less, report more mobility difficulties, and are less likely to be smokers than the normal group. Obesity related health shocks are reported 54 percent more often amongst the obese, while the non-obesity related events occur with similar frequency between groups.

Table 4: Summary Statistics by BMI Group		
	Normal	**Obese**
Bayesian		
P75	0.6919 [0.2383]	0.6624 [0.2272]
P75_lag	0.6981 [0.2282]	0.6648 [0.2488]
X		
education	13.12 [2.860]	12.78 [2.852]
lnassets	12.3100 [1.45]	12.1900 [1.350]
male	0.5748 [0.4946]	0.7794 [0.4147]
female	0.4252 [0.4946]	0.2206 [0.4147]
white	0.9252 [0.2632]	0.8863 [0.3175]
black	0.0544 [0.2269]	0.0887 [0.2844]
hispanic	0.0306 [0.1723]	0.0569 [0.2316]
other	0.0204 [0.1415]	0.0250 [0.1562]
work	0.6037 [0.4893]	0.6088 [0.4881]
unemployed	0.1156 [0.3199]	0.0792 [0.2701]
retired	0.2806 [0.4495]	0.3120 [0.4634]
married	0.9872 [0.1123]	0.9875 [0.1112]
not married	0.0128 [0.1123]	0.0125 [0.1112]
northeast	0.1412 [0.3483]	0.1569 [0.3638]
midwest	0.2500 [0.4332]	0.3025 [0.4594]
west	0.1915 [0.3928]	0.1554 [0.3624]
south	0.4184 [0.4935]	0.3851 [0.4867]

Table 4 Cont.: Summary Statistics by BMI Group		
	Normal	**Obese**
H		
BMI	23.47 [2.09]	29.60 [4.328]
BMI_lag	23.26 [1.969]	29.41 [4.164]
exercise	0.6037 [0.4893]	0.5254 [0.4994]
smoke now	0.1879 [0.3908]	0.1304 [0.3368]
drink	0.8041 [1.195]	0.8067 [1.359]
mobility	0.3937 [0.9108]	0.6478 [1.162]
large muscle	0.6706 [1.108]	0.9052 [1.184]
OBESE		
heart angina	0.0196 [0.1385]	0.0349 [0.1835]
heart attack	0.0128 [0.1123]	0.0201 [0.1403]
conjestive heart failure	0.0034 [0.0582]	0.0095 [0.0969]
diabetes, insulin	0.0009 [0.0292]	0.0008 [0.0275]
diabetes, medication	0.0051 [0.0713]	0.0174 [0.1309]
arthritis	0.0102 [0.1005]	0.0197 [0.1390]
high blood pressure	0.0383 [0.1919]	0.0413 [0.1991]
stroke, medication	0.0060 [0.0770]	0.0045 [0.0673]
stroke, problem	0.0034 [0.0582]	0.0049 [0.0700]
obese shock	0.0995 [0.3493]	0.1532 [0.4542]
NON-OBESE		
cancer	0.0034 [0.0582]	0.0030 [0.0550]
lung disease	0.0051 [0.0713]	0.0068 [0.0823]
non-obese shock	0.0085 [0.0919]	0.0099 [0.0988]

VII. Bayesian Model

a. Model Specification

The empirical model described in section V.a. appears as follows:

$$P_{it} = \left(\frac{\theta}{\theta+\gamma}\right) \bullet P_{i,t-1} + \sum_{j=1}^{k} \alpha_j x_j + \alpha_i + \varepsilon_{it} \quad (55)$$

where

$P_{i,t-1}$ = one wave lag of the subjective longevity report, or prior

$\left(\frac{\theta}{\theta+\gamma}\right)$ = relative precision weight of the prior report, or marginal impact

x_j = risk factors, including health shocks and all other forms of information

α_j = relative precision of the risk factors, or marginal impact of jth information

α_i = unobserved effect

ε_{it} = error term

Equation 53 can be more specifically expressed for estimation as:

$$P75_{it} = \beta_1 P75_{i,t-1} + \beta_2 X_{it} + \beta_3 H_{i,t} + \alpha_i + \varepsilon_{it} \quad (56)$$

where

$P75_{it}$ = subjective probability of living to age 75, as reported for individual i in wave t

$\beta_1 = \left(\frac{\theta}{\theta+\gamma}\right)$ = relative precision weight of the prior

X_{it} = a vector of socioeconomic/demographic characteristics of individual i in wave t

H_{it} = a vector of the reports of health events since the last interview

α_i = unobserved effect

ε_{it} = error term

The goal of the first estimation model, is to determine to what extent prior beliefs and new information influence the posterior longevity report, and if there is a difference between the obese and non-obese samples. Thus, equation 54 is estimated for both BMI groups, with a particular interest in the resulting estimates of β_1 and β_3. These are the same estimates that Smith et al. discussed in their paper, and correspond to the crucial elements of the Bayesian updating model. Given statistical significance, the sign of these coefficients, and their relative magnitudes, will give insight to the thought processes of the BMI sample groups, just as it did for Smith et al.'s smoking groups.

The Bayesian model expressed in equation 54 requires an estimation technique that will account for both the panel data structure and any unobserved effects that can be captured with that structure. The two commonly employed unobserved effects models (UEM) are the fixed effects and random effects models. The basic UEM can be written as:

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1,2,\ldots,T \quad (57)$$

where x_{it} is 1 x K and can contain observable variables that change across t and not i, variables that change across i and not t, and those that change across i and t. If i indexes individuals, it is commonly called an individual effect, or individual heterogeneity. c_i is an unobserved effects measure. The u_{it} is an idiosyncratic error as it too can change across t and i.

The choice between the fixed and random effects models is hinged upon whether c_i is viewed as a random variable, or a parameter to be estimated. c_i is labeled a random

effect under the former choice, and a fixed effect under the later. Wooldridge[77] writes that this debate is ill advised for micro-econometric panel data applications. In his opinion, the key issue is whether c_i is correlated with the observed explanatory variables, x_{it}, t = 1,2,…,T.

Wooldridge's theory indicates that a random effect is synonymous with zero correlation between c_i and x_{it}: $Cov(x_{it}, c_i) = 0$, t = 1,2,…,T. An even stronger assumption involving the conditional mean is: $E(c_i | x_{i1},..., x_{iT}) = E(c_i)$, which is needed for full justification of statistical inference. A fixed effect describes the situation when $Cov(x_{it}, c_i) \neq 0$, and $E(c_i | x_{i1},..., x_{iT}) \neq E(c_i)$, or that there is arbitrary correlation between the unobserved effect, c_i, and the observed explanatory variables, x_{it}.

The random and fixed effects estimation techniques thus differ upon the integration of the unobserved effect. The random effects models put c_i in the error term:

$$y_{it} = x_{it}\beta + \upsilon_{it} \quad (58)$$

where $E(\upsilon_{it} | x_i) = 0$, t = 1,2,…,T, and $\upsilon_{it} = c_i + u_{it}$. In comparison, the fixed effects model allows c_i to be correlated to x_{it}. Using the terminology of the original UEM, a necessary assumption of this model is the strict exogeneity of the explanatory variables conditional on c_i: $E(u_{it} | x_i, c_i) = 0$, t = 1,2,…,T. The goal is then to estimate β after eliminating the effect of c_i. This illustrates one of the drawbacks of fixed effects estimation: the loss of n degrees of freedom in the transformation process, as the unobserved effect for each individual, c_i, is eliminated.

---

[77] Wooldridge, Jeffery (2002), although the argument was first made by Mundlak (1978)

Several transformations are used in the literature, including first differencing and the within transformation. The later is used in this analysis. The within transformation is achieved by first averaging the UEM equation over t = 1,2,…,T to get the cross sectional equation:

$$\bar{y}_i = \bar{x}_i \beta + c_i + \bar{u}_i \quad (59)$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$, $\bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$, and $\bar{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$. Subtracting (59) from (57) for each t yields the desired transformation:

$$\begin{aligned} & y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i) \\ & or \qquad\qquad (60) \\ & \ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it} \end{aligned}$$

where $\ddot{y}_{it} \equiv y_{it} - \bar{y}_i$, $\ddot{x}_{it} \equiv x_{it} - \bar{x}_i$, and $\ddot{u}_{it} \equiv u_{it} - \bar{u}_i$. This time demeaning process has eliminated the effect of c_i. In the absence of c_i, equation (60) can then be estimated via pooled OLS. The resulting estimates, $\hat{\beta}_{FE}$, have been shown to be both unbiased and consistent estimates of β.

One issue with the time demeaning process used to achieve the within transformation, is that along with the unobserved effect, any explanatory variable that is not time-variant is eliminated. This is troubling if x_i are thought to explain $y_{it.}$ Common examples of such time invariant explanatory variables are race, gender, and date of birth. This makes the random effects model more attractive in principle, since the time constant variables remain in the regression model. Greene (2003)[78] offers a solution to this

[78] Greene, William H. 2003. *Econometric Analysis.* New Jersey: Pearson Education, Inc.

dilemma. He suggests interacting the time invariant variables with measure that does change over t, such as a year dummy. Thus, the time invariant variable is made time variant, and is thus not eliminated in the time demeaning process.

The first step in estimating an UEM is to determine if the random effects or fixed effects model best describes the estimation framework. This decision is a multiple step selection process. First, if the sample can be assumed to be a random draw from a given population, then the choice is still undetermined. Conversely, a non-random sample is best estimated using fixed effects. Given the choice between fixed and random effects is unclear, several test statistics exist, based upon the difference in the estimation coefficients. The key consideration in choosing between the two models at this point is whether c_i and x_{it} are correlated. Hausman (1978) proposed a test based on the difference between the estimates of the two models, $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$. The test uses the fact that fixed effects is consistent when c_i and x_{it} are correlated, while random effects is not. Thus, a statistically significant difference is interpreted as evidence against the random effects model. The Hausman test is a hypothesis test in which:

$$\begin{aligned} H_0 &: \hat{\beta}_{RE} - \hat{\beta}_{FE} = 0 \\ H_1 &: \hat{\beta}_{RE} - \hat{\beta}_{FE} \neq 0 \end{aligned} \quad (61)$$

Most statistical software packages can perform the Hausman test, and it is the test used in this paper to choose between the random and fixed effects model. The test uses a chi-square statistic, with degrees of freedom equal to the number of variables used in the estimation.

Regardless of the choice between random effects and fixed effects, another issue arises when an estimation equation contains a lagged dependent variable, as in equation

56. Correlation likely exists between the lagged dependent variable, $P75_{i,t-1}$, and the error term, ε_{it}. The lagged dependent variable can be defined as:

$$P75_{i,t-1} = \beta_1 P75_{i,t-2} + \beta_2 X_{i,t-1} + \beta_3 H_{i,t-1} + \alpha_i + \varepsilon_{it-1} \quad (62)$$

Substituting (62) into (56) yields:

$$P75_{it} = \beta_1(\beta_1 P75_{i,t-2} + \beta_2 X_{i,t-1} + \beta_3 H_{i,t-1} + \alpha_i + \varepsilon_{i,t-1}) + \beta_2 X_{it} + \beta_3 H_{i,t} + \alpha_i + \varepsilon_{it} \quad (63)$$

A typical assumption is that $E(\varepsilon_{it} | X_{it}, H_{it}, P75_{it-1}, ...P75_{i0}, \alpha_i) = 0$, with all of the dynamics captured by the first lag. If $L_{it} = P75_{i,t-1}$, under the above assumption ε_{it} is uncorrelated with $(L_{it}, L_{i,t-1}, ..., L_{i1})$, but not with $(L_{i,t+1}, ..., L_{iT})$, since $L_{i,t+1} = P75_{it}$. Multiplying both sides of equation (54) by ε_{it}, and taking the expected value generates:

$$\begin{aligned} &E(P75_{it}\varepsilon_{it}) = E\beta_1(P75_{i,t-1}\varepsilon_{it}) + E\beta_2(X_{it}\varepsilon_{it}) + E\beta_3(H_{it}\varepsilon_{it}) + E(\alpha_i\varepsilon_{it}) + E(\varepsilon_{it}^2) \\ &E(P75_{it}\varepsilon_{it}) = E(\varepsilon_{it}^2) = \sigma_\varepsilon^2 > 0 \end{aligned} \quad (64)$$

Thus, the strict exogeneity assumption never holds in the UEM with lagged dependent variables.

One solution is to instrument the lagged dependent variable with a variable that is not correlated with ε_{it}. A valid instrument must meet two criteria. First, the instrument, z_1, must be uncorrelated with ε_{it}: $Cov(z_1, \varepsilon_{it}) = 0$, meaning that it is exogenous in equation (56). Second, z_1 need be partially correlated with $P75_{i,t-1}$, once the other exogenous variables have been netted out. If $P75_{i,t-1}=x_K$, and all other explanatory variables are represented by $(x_1,\ldots x_{K-1})$, then the linear projection of x_K onto all of the exogenous variables is:

$$x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K \quad (65)$$

or, more specifically, here (65) can be written as:

$$P75_{i,t-1} = \delta_0 + \delta_1 X_{it} + \delta_2 H_{it} + \theta_1 z_1 + r_K \quad (66)$$

where, by definition of a linear projection error, $E(r_K) = 0$ and r_K is uncorrelated with x_1, $x_2,\ldots,z_1$. The key assumption of the linear projection above is that the coefficient associated with the instrument is non-zero: $\theta_1 \neq 0$. If x_K is the only explanatory variable in the estimation equation then this is equivalent to: $Cov(z_1, x_K) \neq 0$. If these two criteria are met, then z_1 is considered an instrumental variable candidate for x_K. Because $(x_1,\ldots x_{K-1})$ are already uncorrelated with ε_{it}, they serve as their own instrumental variables in the estimation equation (56).

Equation (65) is referred to as a reduced form equation for the endogenous explanatory variable x_K, or here $P75_{i,t-1}$. Substituting the reduced form equation for x_K into the structural equation (56) generates a reduced form equation for $P75_{it}$:

$$P75_{it} = \phi_0 + \phi_1 X_{it} + \phi_2 H_{it} + \alpha_i + \lambda_1 z_1 + v_{it} \quad (67)$$

where $v_{it} = \varepsilon_{it} + \beta r_K$ is the reduced form error, $\phi_j = \beta_j + \beta_K \delta_j$, and $\lambda_1 = \beta_K \theta_1$. According to the criteria for z_1, v_{it} is uncorrelated with all of the explanatory variables in (67), and the reduced form coefficients are estimated consistently.

The instrumental variable approach is appealing, but locating functional instruments within the dataset often proves difficult. The HRS dataset contains several instrumental variable candidates. Anderson and Hsaio (1982)[79] suggest using an additional lag or the difference of the additional lag of the dependent variable as an

---

[79] Anderson, T.W., Cheng Hsaio. (1982). "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.

instrument for the single lag or its difference. More specifically, they suggest using y_{i2} or $(y_{i3}-y_{i2})$ as instruments for y_{i1} or $(y_{i2}-y_{i1})$. It is reasonable to assume that $P75_{i,t-2}$ is both uncorrelated with ε_{it} and partially correlated with $P75_{i,t-1}$. These assumptions can be tested. The HRS dataset used in this paper contains five waves of data, so the Anderson-Hsiao instrument is possible. One drawback to using the double lag of the dependent variable as an instrument for the single lag is that a wave of data is lost. If five waves of data are available, the earliest the estimations can be executed are in wave three, using $P75_{i1}$ as an instrument for $P75_{i2}$ in the estimation of $P75_{i3}$.

Denes-Raj and Ehrlichman[80] report that individuals who lost a parent prematurely estimated that their own lives would be shorter than those of other individuals in the same age-cohort. Individual subjective life expectancy report was significantly lower than those of the same age who did not lose a parent prematurely. This trend suggests that an indicator for premature death of a respondent's father and/or mother is an instrumental variable candidate for $P75_{i,t-1}$. The HRS dataset includes information about the age of each parent at death, which can be used to construct an indicator of premature death for each parent. Following Denes-Raj and Ehrlichman, premature death is established at age 55 or younger. Given these indicator variables meet the two criteria for instrumental variables discussed above, the set of instruments is increased by two, and includes:

$$z_1 = \left(P75_{i,t-2}, Dadpre_{it}, Mompre_{it}\right).$$

[80] Denes-Raj, Veronoka, and Howard Ehrlichman. (1991). "Effects of Premature Parental Death on Subjective Life Expectancy, Death Anxiety, and Health Behavior," *The Journal of Death and Dying* 23(4).

b. Estimation

1. Pooled Ordinary Least Squares

Equation (56) is first estimated via the pooled ordinary least squares (POLS) technique. POLS is generally not considered an optimal estimation method for panel datasets as it fails to accurately accommodate individual effects, specified in equation (56) as α_i. Regardless, it serves as a useful benchmark and the POLS estimates can be compared to others for robustness.

Several requirements exist for valid POLS estimates. These assumptions are:

1. The dependent variable is defined as:
   $$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + ... + \beta_k x_{itk} + a_i + u_{it}$$
2. The matrix of **X** is drawn from a random sample.
3. Each $\mathbf{X_i}$ changes over time.
4. There are no perfect linear relationships between any $\mathbf{X_i}$.
5. $E[u_{it} \mid X_i] = 0$, $E[u_{it} \mid X_i, a_i] = 0$, $E[\Delta u_{it} \mid X_i] = 0$.
6. $Var(\Delta u_{it} \mid X_i) = \sigma^2$
7. $Cov(\Delta u_{it}, \Delta u_{is} \mid X_i) = 0, t \neq s$

The models proposed in this paper are defined according to the first assumption. The HRS dataset is selected to be a nationally representative sample, satisfying the second assumption. Some of the included explanatory variables do not vary over time, including race, gender, region of residence, and years of education, which violates the third assumption. There are two solutions to this problem. First, the time invariant variables can be eliminated from the dataset. This is only a problem if these variables have explanatory power for the dependent variable. The second solution is that these variables can be forced to vary over time by interacting them with a time variant variable, such as

year. Both of these solutions are implemented in the estimation procedure. The forth assumption is satisfied, as there is no perfect collinearity present in the models.

The fifth assumption of a zero conditional mean is always violated in models with lagged dependent variables, such as the models included in this paper. Violation of this assumption yields biased estimates, and any test statistics, and their related hypotheses, are invalid. Additionally, assumption seven is also violated in a model with lagged dependent variables. This assumption assures that the homoskedastic errors are not serially correlated.

Serial correlation, or auto correlation in the lagged dependent variable case, exists if $Corr(u_{it}, u_{is}) \neq 0$ for t not equal to s. In fact, if $u_{i,t-1} > 0, u_{it} > 0$, then $Corr(u_{it}, u_{i,t-1}) > 0$. The interpretation of this outcome is that if y is unexpectedly high in one period, then it is unexpectedly high in the next period. The error, u_{it}, is defined according to the equation: $u_{it} = \rho u_{i,t-1} + e_{it}$. The variance of the estimated coefficient β_j is then $Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_{X_j}} + 2\left(\frac{\sigma^2}{SST_{X_j}}\right)\sum_{t=1}^{n-1}\sum_{m=1}^{n-t}\rho^m x_t x_{t+m}$, where the second term on the right hand side of the equation arises due to the presence of serial correlation amongst the errors. If $\rho > 0$, which is the most common case, then the second term is positive, and the usual estimates underestimate the variance of the estimated coefficients. The standard error is also underestimated, yielding t-statistics that are too large. Additionally, null hypotheses using the test statistics can be rejected when they should not be.

$\hat{\beta}_j$'s are inconsistent if $\rho \neq 0$ since $Cov(y_{i,t-1}, u_{it}) = \rho Cov(y_{i,t-1}, u_{i,t-1}) \neq 0$. This problem can be solved if an additional lag is included in the model. The error form of the

Bayesian model is $y_{it} = \beta_0 + \beta_1 y_{i,t-1} + ... + u_{it} = \beta_0 + \beta_1 y_{i,t-1} + ... + (\rho u_{i,t-1} + e_{it})$.

Assuming $u_{i,t-1} = y_{i,t-1} - \beta_0 - \beta_1 y_{i,t-2} - ...$, the equation becomes

$y_{it} = \beta_0 + \beta_1 y_{i,t-1} + ... + (\rho(y_{i,t-1} - \beta_0 - \beta_1 y_{i,t-2} - ...) + e_{it}) = \alpha_0 + \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + ... + e_{it}$
where $\alpha_0 = (1-\rho)\beta_0, \alpha_1 = \beta_1 + \rho, \alpha_2 = -\rho\beta_1$.

$E[y_{it} | y_{i,t-1}, y_{i,t-2}, ...] = E[y_{it} | y_{i,t-1}, y_{i,t-2}] = \alpha_0 + \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + ...$, yielding consistent estimates of α_j. In sum, if autocorrelation exists in a model with a lagged dependent variable, the serial correlation can be eliminated with the inclusion of a second lagged term. POLS coefficients can be compared to UEM estimates to determine which best describe the normal and obese respondents reaction to adverse health events.

2. UEM Selection

The Hausman test is used to determine if the random effects or fixed effects model is more appropriate when estimating equation (56). The hypothesis test that constitutes the Hausman test is:

$$\begin{aligned} H_0 &: \hat{\beta}_{RE} - \hat{\beta}_{FE} = 0 \\ H_1 &: \hat{\beta}_{RE} - \hat{\beta}_{FE} \neq 0 \end{aligned} \quad (68)$$

To do so, equation 56 is estimated first under the random effects model, and second using the fixed effects model for the obese and normal samples. The coefficients are then tested for a statistically significant difference, using a chi-square statistic. The resulting chi-square statistics are 2731.81 and 1352.62 for the obese and normal samples, respectively. The null hypothesis in favor of the random effects model is soundly rejected in each case. This suggests that the fixed effects model produces more

consistent estimates for both the obese and normal estimation equations and is used to estimate equation (56) for both BMI groups.

Equation (56) is estimated via three methods: POLS, fixed-effects and fixed effects-instrumental variables. The matrix of possible instruments for lagged dependent variable is $z_1 = \left(P75_{i,t-2}, Dadpre_{it}, Mompre_{it}\right)$. The estimation equation (56) becomes:

$$P75_{it} = \phi_0 + \phi_1 X_{it} + \phi_2 H_{it} + \alpha_i + \lambda_1 z_1 + v_{it} \quad (69)$$

The fixed effects estimators, ϕ_j and λ_1, using the within transformation, are written generally as:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{Y}_i\right) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}' \ddot{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}' \ddot{y}_{it}\right) \quad (70)$$

The POLS estimators ϕ_j and λ_1 are written generally as:

$$\hat{\beta}_{POLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}' x_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}' y_{it}\right) \quad (71)$$

The health event indicator variables, H_{it}, are classified as obesity and non-obesity related in the first round of estimations, and in the second, the shocks are included separately. Grouping the health events allows for a greater frequency of occurrence, and thus is more likely to produce significant results. The separation of the events allows further insight into the health events that are driving the results, and the estimation of the separate risk assessments the obese and normal samples associate with each event.

c. Results

1. POLS

Tables 5-8 show the correlation coefficients and covariances, respectively, between the Bayesian dependent variable, and two of its lagged terms for both BMI groups. The relationship between P75 and its lags is positive, but appears to dampen

Table 5: Correlation Coefficients (Normal)

	P75	P75_lag	P75_lag2
P75	1.00		
P75_lag	0.4777	1.00	
P75_lag2	0.4414	0.4347	1.00

Table 6: Covariances (Normal)

	P75	P75_lag	P75_lag2
P75	0.0472		
P75_lag	0.0237	0.0521	
P75_lag2	0.0228	0.0236	0.0568

Table 7: Correlation Coefficients (Obese)

	P75	P75_lag	P75_lag2
P75	1.00		
P75_lag	0.4901	1.00	
P75_lag2	0.4595	0.5082	1.00

Table 8: Covariances (Obese)

	P75	P75_lag	P75_lag2
P75	0.0516		
P75_lag	0.0277	0.0619	
P75_lag2	0.0268	0.0325	0.0659

over time. This suggests that previous longevity beliefs create optimism today, but that the influence of previous beliefs on today's report diminishes over time. Given the second lag captures the time effects of previous beliefs, when included in the POLS model, the variable eliminates the autocorrelation inherent in the lagged dependent regression model. The coefficient estimates from the POLS model with two lags are consistent, and produce valid test statistics.

The POLS model used to estimate the Bayesian model is:

$$P75_{it} = \beta_0 + \beta_1 P75_{i,t-1} + \beta_2 P75_{i,t-2} + \beta_3 X_{it} + \beta_4 H_{it} + u_{it}$$

The individual effect, α_i, is placed in the error term as in the random effects model. The equation is estimated eight times: across both BMI groups, with separate and grouped

health shocks, including both one and two Bayesian lags, allowing for comparisons across BMI groups and tests for relative autocorrelation. The estimation results are shown in Table 10 on page 109. First, the models with one lag are examined, with grouped and separate shocks, followed by an investigation of the estimations with the second lag are investigated. Finally, an analysis of the errors and the level of serial correlation are made, and the "best" POLS model is selected based upon minimizing the correlation of the error terms over time.

Columns [1], [2], [5], and [6] of Table 10 show the POLS estimates of the model with one lagged dependent variable and separate and grouped shocks, respectively, for the obese and normal BMI groups. In all cases, the lagged dependent variable is statistically significant and positive, suggesting that prior beliefs reinforce current beliefs, and create optimism. Little difference exists between BMI groups and are not responsive to grouped versus separate shocks.

Normal blacks are more optimistic than their white counterparts, while obese Hispanics and other minority groups are less positive about the future than their white counterparts. Unemployment engenders pessimism amongst the obese group, but has no impact on the normal sample. Obese and normal respondents are more certain about living to age 75 the greater is their level of education, while men are less so compared to their female peers.

Exercising three of more times a week increases the longevity perceptions of the obese and normal groups by close to two percentage points. Smoking only influences longevity perceptions for the obese. Obese seem to understand the dangers of smoking, as it reduces their longevity report by 5 percentage points. Limitations in mobility reduce

longevity perceptions, especially for normal respondents, who report a partial effect twice that of the obese.

The obese respond to obesity and non-obesity related health events. Diabetes and lung disease reduce the longevity report in the separate shock model, while in the grouped shock model, only non-obesity related health shocks are significant. The negative effect of the significant shocks is remarkable. Diabetes drive down the probability of living to age 75 by 30 percentage points, while lung disease lowers it by 15 percentage points. The normal group only responds to obesity related health events. Heart angina and arthritis are significant and negative amongst the normal sample, both reducing the reported probability by 9 percentage points.

Lagged dependent variables are vehicles for autocorrelation, resulting in invalid test statistics. Tables 5-8 show correlation exists over time between the subjective longevity reports. Tests exist to determine if this correlation is present in the POLS estimation results. The residuals from the POLS models are regressed on their lagged value according to the equation: $u_{it} = \rho u_{i,t-1} + e_{it}$. If $\rho \neq 0$, autocorrelation exists, thus if the null hypothesis $H_0 : \rho = 0$ is rejected in favor of the alternative $H_a : \rho \neq 0$ the estimated coefficients from POLS are invalid and corrections must be made if they are to be used in analysis. Columns [1] and [3] in Table 9 contain the estimates from the regression of the POLS residuals for the single lag model of both BMI groups on its lagged value. The null hypothesis that rho is equal to zero is rejected for both the normal and obese groups, suggesting that autocorrelation exists in the single lagged model.

	Table 9: OLS Estimates of $u_{it} = \alpha + \rho u_{i,t-1} + e_{it}$, POLS Models			
	Normal		Obese	
	(1) Single Lag	(2) Double Lag	(3) Single Lag	(4) Double Lag
α	0.002 [0.971]	0.0003 [0.954]	-0.0001 [0.987]	-0.00004 [0.990]
$u_{i,t-1}$	0.0731 [0.013]	0.0431 [0.142]	0.0359 [0.065]	0.0229 [0.241]
R^2	0.0053	0.0018	0.0013	0.0005
SST	40.70	37.56	99.18	92.38
SSE	0.2151	0.0694	0.1278	0.0482
SSR	40.49	37.49	99.05	92.33
MSE	0.0346	0.0321	0.0376	0.0350
RMSE	0.1861	0.1791	0.1939	0.1872

Wooldridge (2004) suggests incorporating a second lag into the model to capture the variability in the dependent variable explained by its past values as a solution to the autocorrelation problem inherent in models with lagged dependent variables. Columns [3], [4], [7], and [8] in Table 10 display the POLS estimates of the Bayesian model with two lagged dependent variables for the obese and normal groups, respectively. Inclusion of the second lag reduces the effect of the first lagged variable for both BMI groups by approximately ten percentage points. The coefficients remain positive and statistically significant, indicating that the double lag model describes behavior of Bayesian updaters. Together the lags explain close to 50 percentage points of the subjective longevity report for the obese and normal respondents. The second lag has a smaller effect than the first lag, suggesting that the Bayesian effect is diminishing over time.

There is little change in the socioeconomic and demographic estimated coefficients from the single lag to double lag model. Some changes do appear amongst the health variables in matrix H_{it}, some differing by BMI group. In the single lag model,

mobility has a negative and significant impact on the longevity perceptions of the obese and normal groups, although the effect is greater for the normal group. The addition of the second lag increases the gap between the normal and obese response to mobility difficulties. Normal respondents reduce their longevity report by three times the amount obese respondents do after an additional mobility difficulty in the double lag model. Obese remain the only group to respond to the dangers of smoking, although the effect is diminished by almost 20% with the addition of the second lag.

Responses to health shock variables are similar to those of the single lag model. The obese respond to the same obesity and non-obesity related health events, diabetes and lung disease. The coefficient for diabetes increases to over -0.315, while that of long disease falls to –0.137, both being statistically significant. The obese are increasingly responsive to the obesity related health event, and less responsive to the non-obesity related complication in the double lag model. In the grouped shocks model, the obese are remain responsive only to non-obesity related episodes, and this response is similar for the single and double lagged models. The normal group responds only to angina in the double lagged case, as the coefficient on arthritis is no longer significant. The response to angina remains negative, and is slightly smaller in the double lag model.

The double lag model suggests that the obese respond to both obesity related and non-obesity related health events, while the normal group responds only to angina, an obesity related health shock. These results are the same as in the single lag model, save the insignificance of arthritis in the normal case. Past beliefs influence current beliefs for both BMI groups, although the effect appears to be spread across time and the inclusion of a second lag captures the diminishing impact over time. The addition of the second

lagged term increases the model F-statistic, estimated sum of squares and model R^2 as expected.

The purpose of including a second lagged term is to eliminate the autocorrelation inherent in models with lagged dependent variables. The residuals from the POLS estimation of the double lagged model are regressed upon their lagged value to determine if the autocorrelation existent in the single lag model is eliminated with the inclusion of the second lagged term. OLS is used to estimate the equation $u_{it} = \alpha + \rho u_{i,t-1} + e_{it}$, and the results are shown in columns [2] and [4] of Table 9. The null hypothesis $H_0 : \rho = 0$ is tested against the alternative $H_a : \rho \neq 0$. If $\rho \neq 0$, autocorrelation exists, and the inclusion of the second lagged term did not solve the autocorrelation issue. The p-values associated with rho are 0.142 and 0.241 for the normal and obese groups, respectively. The null hypothesis cannot be rejected for either BMI group indicating that the addition of the second lagged term successfully eliminated the autocorrelation. This is significant, as the double lag model yields valid test statistics, and inference can be made regarding the model's estimated coefficients.

Table 10: Pooled OLS Estimates for Obese and Normal BMI Groups

	Obese				Normal			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	Pooled OLS	Pooled OLS Grouped	Pooled OLS (2)	Pooled OLS (2) Grouped	Pooled OLS	Pooled OLS Grouped	Pooled OLS (2)	Pooled OLS (2) Grouped
Bayesian								
P75_lag	0.3994	0.4020	0.2890	0.2901	0.4022	0.4025	0.3048	0.3041
	[24.69]	[25.01]	[16.45]	[16.60]	[15.85]	[15.89]	[11.53]	[11.54]
P75_lag2			0.2321	0.2329			0.2410	0.2443
			[13.75]	[13.89]			[9.62]	[9.77]
X								
education	0.0032	0.0034	0.0220	0.0024	0.0067	0.0068	0.0060	0.0062
	[2.21]	[2.35]	[1.58]	[1.74]	[3.14]	[3.25]	[2.94]	[3.06]
assets	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[1.49]	[1.50]	[1.14]	[1.14]	[0.45]	[0.44]	[0.59]	[0.58]
male	-0.0166	-0.0167	-0.0124	-0.0124	-0.0287	-0.0293	-0.0249	-0.0248
	[-1.65]	[-1.66]	[-1.28]	[-1.29]	[-2.36]	[-2.42]	[-2.13]	[-2.13]
black	0.0183	0.0175	0.0172	0.0162	0.0545	0.0560	0.0482	0.0496
	[1.34]	[1.29]	[1.30]	[1.24]	[2.16]	[2.24]	[1.98]	[2.06]
hispanic	-0.0303	-0.0297	-0.0288	-0.0280	-0.0009	0.0046	0.0222	0.0280
	[-1.69]	[-1.66]	[-1.67]	[-1.63]	[-0.03]	[0.13]	[0.66]	[0.85]
other	-0.0529	-0.0535	-0.0556	-0.0557	-0.0476	-0.0435	-0.0348	-0.0299
	[-2.09]	[-2.12]	[-2.28]	[-2.29]	[-1.21]	[-1.10]	[-0.92]	[-0.79]
unemployed	-0.0287	-0.0283	-0.0254	-0.0245	-0.0012	-0.0018	0.0076	0.0079
	[-1.84]	[-1.82]	[-1.69]	[-1.63]	[-0.07]	[-0.10]	[0.42]	[0.44]
retired	0.0042	0.0045	0.0009	0.0014	0.0009	-0.0003	0.0049	0.0038
	[0.50]	[0.52]	[0.10]	[0.17]	[0.07]	[-0.02]	[0.40]	[0.31]
married	0.0540	0.0545	0.0552	0.0560	-0.0286	-0.0230	-0.0267	-0.0236
	[1.57]	[1.58]	[1.66]	[1.68]	[-0.57]	[-0.46]	[-0.55]	[-0.49]
northeast	-0.0242	-0.0241	-0.0233	-0.0232	-0.0123	-0.0109	-0.0118	-0.0102
	[-1.74]	[-1.74]	[-1.73]	[-1.73]	[-0.63]	[-0.55]	[-0.62]	[-0.54]
midwest	-0.0175	-0.0169	-0.0167	-0.0158	-0.0253	-0.0230	-0.0181	-0.0151
	[-1.42]	[-1.37]	[-1.40]	[-1.33]	[-1.44]	[-1.32]	[-1.07]	[-0.90]
south	-0.0191	-0.0184	-0.0164	-0.0157	-0.0098	-0.0090	-0.0051	-0.0041
	[-1.62]	[-1.58]	[-1.45]	[-1.38]	[-0.61]	[-0.56]	[-0.33]	[-0.26]
H								
exercise	0.0189	0.0179	0.0155	0.0146	0.0204	0.0216	0.0171	0.0185
	[2.38]	[2.25]	[2.01]	[1.89]	[1.75]	[1.86]	[1.52]	[1.66]
smokes now	-0.0518	-0.0520	-0.0136	-0.0436	-0.0012	-0.0002	0.0043	0.0055
	[-4.45]	[-4.48]	[-3.87]	[-3.88]	[-0.08]	[-0.01]	[0.30]	[0.38]
drink	0.0029	0.0028	0.0029	0.0028	-0.0009	-0.0011	-0.0011	-0.0015
	[1.02]	[0.98]	[1.04]	[1.01]	[-0.18]	[-0.23]	[-0.24]	[-0.31]
mobility	-0.0132	-0.0128	-0.0082	-0.0077	-0.0272	-0.0267	-0.0221	-0.0214
	[-3.55]	[-3.50]	[-2.27]	[-2.18]	[-3.94]	[-3.92]	[-3.31]	[-3.27]
OBESE								
heart angina	-0.0192		0.0021		-0.0911		-0.0847	
	[-0.81]		[0.09]		[-1.97]		[-1.91]	
heart attack	-0.0475		-0.0459		0.0253		0.0233	
	[-1.58]		[-1.59]		[0.49]		[0.46]	
congestive heart failure	0.0083		-0.0060		0.0815		0.0796	
	[0.20]		[-0.15]		[0.80]		[0.91]	
diabetes, insulin	-0.3022		-0.3148		0.0577		0.0954	
	[-2.17]		[-2.34]		[0.27]		[0.47]	
diabetes, medication	0.0202		0.0083		0.0436		0.0090	
	[0.69]		[0.30]		[0.50]		[0.11]	
arthritis	0.0214		0.0286		-0.0902		-0.0803	
	[0.78]		[1.08]		[-1.64]		[-1.51]	
high blood pressure	0.0102		0.0024		-0.0085		-0.0014	
	[0.53]		[0.13]		[-0.29]		[-0.05]	
stroke, medication	0.0589		0.0461		-0.1042		-0.0709	
	[0.86]		[0.70]		[-1.16]		[-0.82]	
stroke, problem	-0.0391		-0.0566		0.1643		0.1025	
	[-0.59]		[-0.89]		[1.42]		[0.92]	
obese shock		-0.0096		-0.0083		-0.0177		-0.0144
		[-1.16]		[-1.04]		[-1.16]		[-0.98]
NON-OBESE								-0.0144
lung disease	-0.1524		-0.1367		0.0968		0.0795	[-0.98]
	[-3.27]		[-3.04]		[1.21]		[1.03]	
non-obese shock		-0.1521		-0.1525		0.0002		-0.0276
		[-3.94]		[-4.10]		[0.00]		[-0.47]
N	2638	2638	2638	2638	1174	1174	1174	1174
F	36.20	51.29	44.18	61.95	15.74	21.86	19.70	27.24
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R-square	0.2725	0.2713	0.3217	0.3213	0.2705	0.2647	0.3251	0.3209
Adj R-square	0.2649	0.2660	0.3144	0.3161	0.2533	0.2526	0.3086	0.3091
SSE	37.09	36.93	43.79	43.74	14.99	14.66	18.01	17.78
SSR	99.04	99.21	92.35	92.39	40.41	40.74	37.39	37.62
SST	136.14	136.14	136.14	136.14	55.40	55.40	55.40	55.40
MSE	0.0379	0.0389	0.0354	0.0353	0.0353	0.0353	0.0327	0.0326
Root MSE	0.1948	0.1947	0.1881	0.1879	0.1878	0.1879	0.1807	0.1806

2. Fixed Effects

A UEM is often preferable to POLS when analyzing panel data. Consider again the linear unobserved effects model, $y_{it} = x_{it}\beta + \alpha_i + u_{it}$. The POLS approach puts the individual effect, α_i, in the error term, just as random effects method does. This is undesirable if the purpose of using a panel dataset is to allow the individual effect to be arbitrarily correlated with the x_{it}. The optimal model to understand the Bayesian and behavioral response of aging Americans to obesity related health events should explicitly estimate the individual effect and allow it to be any function of x_{it} if those individual effects exist. The Hausman test conducted in section b.2. of chapter VII suggests that individual effects exist within the Bayesian model, and therefore should not be simply placed in the error term, but allowed to be correlated with x_{it} and be a part of the estimated portion of the equation such that $E[\alpha_i \mid x_i]$ is any function of x_i.

Equation (56), $P75_{it} = \beta_1 P75_{i,t-1} + \beta_2 X_{it} + \beta_3 H_{i,t} + \alpha_i + \varepsilon_{it}$, is estimated via the fixed effects method, and the results are shown in columns [1], [2], [5], and [6] of Table 12 for the obese and normal BMI groups, respectively. Prior beliefs create pessimism for both the obese and normal respondents. The coefficients for the lagged dependent variables are significant and negative for both BMI groups. The values are similar in magnitude to the POLS estimates, but of the opposite sign. These results should be interpreted with caution, however, as the presence of autocorrelation leads to invalid test statistics. Substantial correlation exists between the residuals, u_{it}, and the explanatory variables. Table 12 shows that corr(u_it, Xb) is –0.7764 and –0.7432 for the obese and normal groups, respectively. Minor correlation between explanatory variables and the model error is not a major concern, but that reported in the fixed effects models is

substantial. Correlations are on the [-1,1] range, with the extremes signifying perfect linear relationships. The correlations in the fixed effects model are too large and too close to –1.0 to be ignored.

Columns [1] and [3] in Table 11 display the coefficients from the regression of the fixed effects residuals, u_{it}, on $u_{i,t-1}$. The coefficients are close to 0.5 for both the obese and normal groups. Wooldridge (2002) suggests using these coefficients, along with their standard errors, to test the hypothesis $H_0 : \rho = -1/(T-1)$, where rho is the correlation between u_{it} and $u_{i,t-1}$. This hypothesis is rejected for both the obese and normal groups, suggesting that there is not a perfect negative linear relationship. The p-values shown in Table 11 are from the hypothesis $H_0 : \rho = 0$. This hypothesis is soundly rejected for both BMI groups. Finally, the hypothesis that the errors follow an AR(1) process, $H_0 : \rho = -0.5$ is cannot be rejected for either BMI groups even at the 90th percentile. Thus, there appears to be some serial correlation present, perhaps best described by an AR(1) process, or $u_{it} = -0.5u_{i,t-1} + e_{it}$. Regardless of the form of serial correlation present in the fixed effects model, its mere presence renders the FE coefficient dubious at best.

	Table 11: OLS Estimates of $u_{it} = \alpha + \rho u_{i,t-1} + e_{it}$, FE Models			
	Normal		Obese	
	(1) FE	(2) FE-IV	(3) FE	(4) FE-IV
α	0.0000 [0.999]	0.0000 [0.998]	-0.0001 [0.984]	-0.0001 [0.985]
$u_{i,t-1}$	0.5427 [0.000]	0.5428 [0.000]	0.5309 [0.000]	0.5310 [0.000]
R^2	0.2935	0.2937	0.2820	0.2820
SST	42.30	45.35	102.21	111.27
SSE	12.42	13.32	28.82	31.38
SSR	29.89	32.03	73.39	79.88
MSE	0.0256	0.0274	0.0279	0.0303
RMSE	0.1599	0.1655	0.1669	0.1741

Time invariant explanatory variables, including gender, race, education, and region of residence are lost in the time demeaning process[81]. The remaining variables in the X_{it} matrix are insignificant for both BMI groups. Several explanatory variables in the H_{it} matrix are significant, however. Drinking creates optimism for the obese sample and pessimism for the normal group. The effect is small in both cases, as are the sample means. The obese respond to normal and obesity related health events. Angina increases the obese respondent's subjective probability of living to age 75. The optimism produced by angina may be a display of relief. The physical symptoms of angina are similar to those of a heart attack, although angina is not in itself life threatening. Thus a diagnosis of angina, rather than a heart attack, could engender optimism. Diabetes has a substantial negative effect on obese respondents longevity perceptions. A diagnosis of diabetes lowers the reported probability by close to 31 percentage points. Lung disease, a non-

[81] The fixed effects approach was used to estimate equation 56, but with the time invariant variables interacted with year so as to render them time variant. None of the interacted variables are significant with this approach, so it is not used for inference.

obesity related complication, also reduces the longevity report, although its effects are much smaller in magnitude than a diabetes diagnosis. Respondents comprising the normal BMI group do not react to either type of health event. Despite these findings, elimination of the autocorrelation present in the fixed effects model will produce more reliable estimates. The troublesome lagged dependent variable is instrumented in the following section in attempt to do just that.

There is one issue that needs to be addressed before turning to instrumentation as a solution to the autocorrelation problem. The premise behind Bayesian updating is that individuals use all available information, including prior beliefs, when forming their current longevity perceptions. If individuals are Bayesian updaters, then it must be that prior longevity reports are correlated with the current report. Instrumenting for prior beliefs reduces this relationship in an attempt to produce valid test statistics, but in the process may remove just the behavior the model is attempting to capture. The solutions to the autocorrelation problem in this dissertation preserve the relationship between the posterior and prior beliefs. The POLS model uses an additional lag of the subjective probability report, and the instrument used in the FE-IV estimation approach uses a double lag as an instrument for the first. It is imperative to minimize the autocorrelation inherent in models with lagged dependent variables, but this should be done in a manner that preserves the Bayesian relationship the model is structured to capture.

3. Fixed Effects, Instrumented Variables

Fixed effects with instrumental variable estimates of equation (69) for the Bayesian updating model are shown in Table 12. The obese and normal groups respond

to different types of information when forming their posterior longevity report. Heart attack, lung disease and diabetes decrease the obese group's subjective probability of living to age 75. The coefficient associated with diabetes remains considerable in magnitude. A diagnosis of diabetes lowers the probability report by 30.7 percentage points. Heart attack reduces the obese respondents' subjective probability of living to age 75 by 5.7 percentage points and lung disease by 11 percentage points. These adverse health events are categorized as both obesity and non-obesity related. This is in contrast to Smith et al.'s findings. In their results, the focus group, smokers, responded only to smoking related events. None of the variables in the X_{it} matrix are significant, although the number of drinks consumed per day remains significant and positive.

The normal group only responds to stroke. Those respondents taking medication to treat stroke, are more pessimistic about living to age 75 or more, compared to those reporting mobility difficulties after a stroke, who are more optimistic. Stroke is an obesity related event. Thus, the normal group only appears to respond to obesity related health conditions. Again, this is in contrast to the Smith et al. findings. In that paper, the non-smokers responded to both smoking and non-smoking related health events. It appears that it is not the category of health event that matters, but the type of disease itself when forming longevity expectations. Because each group does not demonstrate significant responses to similar events, a comparison cannot be made as to the relative degree of response to health events by BMI category as Smith et al. were able to do in their work.

Neither the obese nor the normal groups use their prior longevity report when forming their posterior subjective probability in the instrumented version of the fixed

effects estimation approach. The advent of an adverse health event is more influential than prior beliefs when forming current beliefs. This is not to say that these groups are myopic or shortsighted, but rather experiencing and surviving an adverse health event trumps any prior beliefs the respondents have concerning survival to age 75. Whether the influence of these experiences is manifested into a behavioral response remains to be seen in the behavioral model.

Additionally, the autocorrelation present in the fixed effects model is significantly reduced once the lagged dependent variable is instrumented. Rho is reduced from 0.7858 to 0.6269 for the obese and from 0.7829 to 0.6339 for the normal BMI group. Further, the correlation between the residual, u_{it}, and the explanatory variable is reduced from –0.7764 to –0.3651 for the obese and from –0.7432 to –0.2741 for the normal BMI group. Thus, the instrumented version of the fixed effects approach appears to handle the autocorrelation endemic in a model with lagged dependent variables better than fixed effects alone.

	Table 12: Fixed Effects Estimates for Obese and Normal BMI Groups							
	Obese				Normal			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	FE	FE Grouped	FE-IV	FE-IV Grouped	FE	FE Grouped	FE-IV	FE-IV Grouped
Bayesian								
P75_lag	-0.4241	0.4222	-0.0490	-0.0518	-0.4161	0.4219	-0.0446	-0.0654
	[-18.89]	[-18.84]	[0.94]	[-1.01]	[-12.37]	[-12.69]	[-0.60]	[-0.90]
X								
assets	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[-0.40]	[-0.49]	[-0.36]	[-0.41]	[0.20]	[0.18]	[0.38]	[0.35]
unemployed	-0.0003	0.0012	-0.0100	-0.0079	-0.0253	-0.0264	-0.0340	-0.0373
	[-0.01]	[0.05]	[-0.39]	[-0.31]	[-0.78]	[-0.82]	[-0.95]	[-1.06]
retired	-0.0029	-0.0036	0.0043	0.0026	-0.0020	-0.0067	0.0053	-0.0034
	[-0.02]	[-0.25]	[0.27]	[0.16]	[-0.10]	[-0.33]	[0.23]	[-0.15]
married	0.0161	0.0194	0.0804	0.0830	0.0163	0.0182	0.0203	0.0241
	[0.18]	[0.21]	[0.79]	[0.82]	[0.08]	[0.09]	[0.10]	[0.11]
H								
exercise	-0.0016	-0.0021	0.0041	0.0039	-0.0012	-0.0067	-0.0028	-0.0005
	[-0.15]	[-0.19]	[0.35]	[0.33]	[-0.08]	[0.04]	[-0.17]	[-0.03]
smokes now	0.0008	0.0009	0.0052	0.0064	0.0422	0.0380	0.0638	0.0565
	[0.03]	[0.03]	[0.14]	[0.860]	[1.19]	[1.09]	[1.63]	[1.47]
drink	0.0121	0.0124	0.0118	0.0119	-0.0177	-0.0190	-0.0235	-0.0250
	[2.19]	[2.24]	[1.93]	[1.96]	[-2.02]	[-2.20]	[-2.42]	[-2.63]
mobility	-0.0034	-0.0029	-0.0022	-0.0010	-0.0102	-0.0100	-0.0161	-0.0143
	[-0.57]	[-0.49]	[-0.34]	[-0.15]	[-0.97]	[-0.97]	[-1.37]	[1.26]
OBESE								
heart angina	0.0609		0.0539		-0.1081		-0.1346	
	[1.97]		[1.58]		[-1.23]		[-1.39]	
heart attack	-0.0486		-0.0574		0.0130		0.0399	
	[-1.52]		[-1.63]		[0.21]		[0.60]	
congestive heart failure	0.0156		0.0838		0.0014		0.0119	
	[0.31]		[1.49]		[0.21]		[0.08]	
diabetes, insulin	-0.3089		-0.3073		0.0900		-0.0275	
	[-2.12]		[-1.92]		[0.42]		[-0.12]	
diabetes, medication	-0.0439		-0.0170		-0.0134		0.0479	
	[-1.44]		[-0.51]		[-0.15]		[0.48]	
arthritis	0.0242		0.0156		-0.0323		-0.0631	
	[0.85]		[0.50]		[-0.57]		[-1.01]	
high blood pressure	0.0182		0.0211		-0.0101		0.0006	
	[0.92]		[0.97]		[-0.34]		[0.02]	
stroke, medication	0.0170		-0.0054		-0.1166		-0.1992	
	[0.24]		[-0.07]		[-1.27]		[-1.95]	
stroke, problem	-0.0066		0.0184		0.1531	-0.0114	0.2566	
	[-0.10]		[0.24]		[1.29]	[-0.62]	[1.94]	
obese shock		-0.0004		0.0016				-0.0102
		[-0.04]		[0.15]				[-0.51]
NON-OBESE								
lung disease	-0.0880		-0.1137		0.0469		0.0584	
	[-1.82]		[-2.13]		[0.58]		[0.66]	
non-obese shock		-0.0529		-0.0846		-0.0179		-0.0263
		[-1.31]		[-1.91]		[-0.29]		[-0.39]
N	2638	2638	2638	2638	1174	1174	1174	1174
T	2	2	2	2	2	2	2	2
F	20.22	33.38			9.46	16.05		
Prob > F	0.0000	0.0000			0.0000	0.0000		
Wald chi2			46055.71	45494.89			24911.65	25383.1400
Prob > chi2			0.0000	0.0000			0.0000	0.0000
R-square	0.2184	0.2254	0.0247	0.0024	0.1813	0.1999	0.0004	0.0192
corr(u_i, Xb)	-0.7764	-0.7780	-0.3651	-0.3162	-0.7432	-0.7570	-0.2741	-0.3856
sigma_u	0.2755	0.2742	0.2055	0.2064	0.2578	0.2598	0.1970	0.2008
sigma_e	0.1439	0.1443	0.1586	0.1595	0.1358	0.1353	0.1497	0.1482
rho	0.7858	0.7832	0.6269	0.6261	0.7829	0.7865	0.6339	0.6472
F test all u_i=0	2.64	2.62	1.16	1.14	2.77	2.81	1.28	1.29
Prob > F	0.0000	0.0000	0.0039	0.0089	0.0000	0.0000	0.0018	0.0010

4. Model Selection

Autocorrelation is a major hindrance when estimating a model with lagged dependent variables. The coefficients produced from a model characterized by this violation of the classical linear regression assumption result in unreliable test statistics, and inference cannot be made using them. Solutions exist for the autocorrelation problem. Wooldridge suggests adding an additional lag to the POLS model, as the majority of the variation in the dependent variable is typically captured by the combination of the first two lags. This appears to be the case in the Bayesian model. Thus, the best model between the two POLS models is the one with the two lagged subjective probability variables.

Panel data set offer benefits over cross sectional data and time series data sets. Tracing individuals over time allows the econometrician to capture differences between individuals over time, or individual effects. POLS does not control for or capture these individual effects. The Hausman test concluded that individual effects exist in the Bayesian model, and that the fixed effects model is the optimal UEM. The fixed effects estimation approach does not correct for autocorrelation already found to be present in the model. Instrumenting for the lagged dependent variable is a solution to this problem. Test statistics show that the correlation between the residuals and the dependent variables is minimized with fixed effects instrumental variable approach.

A choice must be made between the POLS model with the double lag and the fixed effects instrumental variable approach as the optimal framework for determining if the obese are Bayesian updaters. The model that minimizes the autocorrelation is the best model of the two choices. One key to this decision is the magnitude of rho. Rho is

the coefficient on the lag of the residual in the equation $u_{it} = \rho u_{i,t-1} + e_{it}$, and is important in deciphering the level of autocorrelation present in an estimation model. The variance of the estimated coefficient in the POLS and fixed effects model is

$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_{X_j}} + 2\left(\frac{\sigma^2}{SST_{X_j}}\right)\sum_{t=1}^{n-1}\sum_{m=1}^{n-t}\rho^m x_t x_{t+m}$, where the second term on the right hand side of the equation, including rho, arises due to the serial correlation amongst the errors. If rho is non-zero, autocorrelation exists, and the larger its absolute value, the greater the bias that it creates.

An additional consideration must be the magnitude of the individual effects present within the model. POLS does not accommodate these effects, and if they are substantial, the fixed effects approach may prove superior. To evaluate the size of the individual effect, the random effects model of the error is used. The individual effect, α_i, is placed in the error term, such that $v_{it} = \alpha_i + u_{it}$. Thus the variance of v_{it} is the sum of the variances of α_i and u_{it}, or $\sigma_v^2 = \sigma_\alpha^2 + \sigma_u^2$. The absence of an individual effect is statistically equivalent to $H_0 : \sigma_\alpha^2 = 0$ [82], where

$\hat{\sigma}_\alpha^2 = \left(\frac{1}{[NT(T-1)/2 - K]}\right)\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{\hat{v}}_{it}\hat{\hat{v}}_{is}$, and $\hat{\hat{v}}_{it}$ are the POLS residuals. Table 13 shows estimates of the individual effects error, σ_α^2. Individual effects are present in the model, as the Hausman test suggested, but the individual effects error is minimized in the instrumented version of the fixed effects approach. POLS does not accommodate these effects, suggesting FE is the superior estimation approach. Given the FE-IV approach

[82] Wooldridge (2002), pg. 264.

minimizes the autocorrelation amongst the fixed effects approaches, that is the model suggested for inference concerning the differences between the BMI groups' subjective longevity responses to adverse health events.

	Table 13: Estimates of $\sigma_v^2 = \sigma_\alpha^2 + \sigma_u^2$			
	Normal		**Obese**	
	(1) FE	**(2) FE-IV**	**(3) FE**	**(4) FE-IV**
σ_α^2	0.0665	0.0388	0.0759	0.0422
σ_u^2	0.0089	0.0108	0.0102	0.0124

5. Response Differences

The results indicate that the obese and normal groups respond differently to adverse health events. Two tests are conducted to understand why these differences exist. The first examination tests the equality of the risk equivalents of the obesity related health events across BMI groups, with an aim to determine if the obese view the obesity related health events differently than the non-obese: i.e. do the obese find obesity related events to be more or less risky than the normal respondents do? The estimated coefficient for the obesity related health event from equation (69), $\hat{\phi}_{OS}$, is normalized by one minus the weight assigned to the prior longevity report for each BMI group:

$$\frac{\hat{\phi}_{OS}^{BMI}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} \quad (71)$$

A hypothesis test for equality amongst the normalized risk measures is conducted such that:

$$H_0 : \frac{\hat{\phi}_{OS}^{O}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} - \frac{\hat{\phi}_{OS}^{N}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} \leq 0$$
$$H_1 : \frac{\hat{\phi}_{OS}^{O}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} - \frac{\hat{\phi}_{OS}^{N}}{(1-(\hat{\theta}/\hat{\theta}+\hat{\gamma}))} > 0 \quad (72)$$

where O = obese risk measure, and N = normal risk measure.

A z score of 2.92 rejects the null hypothesis at $\alpha = 0.01$ level. This suggests that the obese assign a higher risk assessment to the obesity related health events that the normal sample does. Thus, the obese respond to a greater degree to obesity related health events. This is the same result that Smith et al. found in their comparison of smoking groups risk assessments. They found that smokers assign a significantly greater weight to smoking related health events than former smokers. The difference was too close to call between current smokers and never-smokers however.

A second test uses a reduced form expression for the risk equivalent of the new information that is assumed communicated through the adverse health events. This risk equivalent is calculated at the mean, such that the reduced form expression becomes:

$$\bar{r} = \frac{\alpha_0 + \sum_j \hat{\phi}_j \bar{x}_j}{1-(\hat{\theta}/\hat{\theta}+\hat{\gamma})} \quad (73)$$

where the denominator is one minus the coefficient for the prior longevity perception, and the numerator includes information from all other sources, including the adverse health events. The estimated updating model, captured in the estimated coefficients, is applied to the mean values of the BMI group's observed experience to estimate the group mean risk equivalent, $\bar{r}$. These estimates allow the effects of each group's weights for information to be separated from their actual health-related experiences, and for comparisons to be made.

Table 14 reports the estimates of $\bar{r}$ by sample. In each case, the sub-sample's estimated updating model is applied to the observed experiences of both sub-samples to estimate $\bar{r}$ for each group as it would be implied by their own and others' use of all new information.

Table 14. Relative Risk Estimates

Sample	Obese	Normal
Obese	0.6351	0.6918
Normal	0.6364	0.6908

The effect of each group's weights for prior information is separated from the actual health experiences, or the new information. For example, the first column applies the obese group's updating model to the respondents from both the obese and normal groups, respectively, for all waves. The estimates indicate that if the obese experienced the health records of the normal, they would still interpret this information about living to age 75 to imply a lower likelihood based upon what they learned than the normal group would. This implied risk equivalent of the new information assigned by the obese is higher than what the normal would assign to their health experiences.

Examining the results across rows indicates the interpretation of the obese to their own experience compared to the normal interpretation to the same experience. The obese do evaluate new information differently then the normal group. Given the obese experience, the obese are more pessimistic about living to age 75, implying a higher risk equivalent for equivalent experiences. The same is true for the normal experience. The normal are again more optimistic about living to age 75 or more. The obese then, despite

the health experience, attach a greater risk equivalent to health information than the normal group does. These two tests yield support for the hypotheses that the obese choose to participate in risky behavior, and interpret new information differently than the normal group. In fact they react more severely than the normal group to new adverse health information. These findings are in line with those obtained in Smith et al.'s examination of smokers, former smokers, and never-smokers. They too found that the riskier group, smokers, interpreted information differently than the other less risky groups, and responded to a greater degree to that information.

VIII. Behavioral Model

a. Model Specification

The empirical model used to estimate the behavioral response of the BMI groups to adverse health events is:

$$BMI_{it} = \left(\frac{\theta}{\theta+\gamma}\right) \bullet BMI_{i,t-1} + \sum_{j=1}^{k} \alpha_j x_j + \alpha_i + \varepsilon_{it} \quad (74)$$

where

BMI_{it} = current BMI report of individual i at time t

$BMI_{i,t-1}$ = lag of BMI of individual i at time t

$\left(\frac{\theta}{\theta+\gamma}\right)$ = relative precision weight of prior BMI report

x_j = risk factors, including health shocks and all other forms of information

α_j = relative precision of the risk factors, or marginal impact of jth information

α_i = unobserved effect

ε_{it} = error term

The matrix X_j contains the same explanatory variables as that of the Bayesian model, save the included health events. Only those health events for which weight loss is shown to improve the quality of life, or reduce the chances of reoccurrence are included[83]. Those health events include heart angina, heart attack, congestive heart failure, arthritis, diabetes, high blood pressure, and stroke.

---

[83] Appendix 7 shows the treatments for each health event. Those health events for which weight loss is recommended are included in the estimation equation.

Equation (74) is more specifically expressed for estimation as:

$$BMI_{it} = \beta_1 BMI_{i,t-1} + \beta_2 X_{it} + \beta_3 H_{i,t} + \alpha_i + \varepsilon_{it} \quad (75)$$

where

BMI_{it} = reported BMI as reported for individual i in wave t

$\beta_1 = \left(\frac{\theta}{\theta + \gamma} \right)$ = relative precision weight of the prior

X_{it} = socioeconomic/demographic characteristics of individual i in wave t

$H_{i,t}$ = report of health events since the last interview

α_i = unobserved effect

ε_{it} = error term

The goal of the behavioral model, is to determine to what extent new information influence current BMI, and if there is a difference between the obese and non-obese samples. Thus, equation (75) is estimated for both groups, with a particular interest in the resulting estimates of β_1 and β_3. Using weight loss as the measure of a behavioral response, this model is used to determine if adverse health information manifests itself in a behavioral response. This is the first model to measure such responses across BMI groups, and is a clear extension of the Bayesian model.

An UEM is needed to estimate the behavioral model, just as it was the Bayesian model presented above to account for the panel data structure of the HRS and the individual effects stemming from it. The behavioral model is similar to the Bayesian model, including the presence of a lagged dependent variable. The estimation issues discussed in Section VII. a. apply. A Hausman test is employed to determine if the behavioral model is best described by the random or fixed effects model. Candidates to

instrument the lagged dependent variable must also be established within the HRS dataset. Anderson and Hsaio's (1982) suggestion of lagging the dependent variable an additional time is selected for this purpose. Thus, $BMI_{i,t-1}$ is instrumented with $BMI_{i,t-2}$ and so on. An additional wave of data will be lost in the process, but the dataset begins with five waves, leaving three waves available using the double-lag instrumenting technique.

b. Estimation

1. UEM Selection

The Hausman test is used to determine if the random or fixed effects model best describes the behavioral model. The hypothesis

$$\begin{aligned} H_0 &: \hat{\beta}_{RE} - \hat{\beta}_{FE} = 0 \\ H_1 &: \hat{\beta}_{RE} - \hat{\beta}_{FE} \neq 0 \end{aligned} \quad (76)$$

is tested for both BMI groups. The chi-square statistic for the obese group is 100.29, rejecting the null hypothesis in favor of a fixed effects model. The hypothesis test of the normal group cannot be rejected, with a chi-square statistic of 23.67. The random effects model best describes the normal behavioral response to health events. The estimates of the normal group in the behavioral model are less interesting than those in the Bayesian model, and a comparison of the obese and normal makes little sense. The normal are a healthy weight by definition. Thus, weight loss post an adverse health event will neither yield improvements in quality of life, nor reduce the probability of reoccurrence of the adverse health event. The obese group is the focal group of this model, and will be

estimated using the fixed effects method and instrumenting the lagged dependent variable with its double lag.

Equation (75) is estimated for the obese group via the fixed effects-instrumental variables model. This procedure is also used for the normal group estimations, but with caution, as the fixed effects model is not the most appropriate method according to the Hausman test. This is of little concern as stated earlier, as the normal group is not the focus of the behavioral model. Using $z_1 = (BMI_{i,t-2})$ as the instrumental variable matrix for the lagged dependent variable, the estimation equation is expressed as:

$$BMI_{it} = \phi_0 + \phi_1 X_{it} + \phi_2 H_{it} + \alpha_i + \lambda_1 z_1 + v_{it} \quad (77)$$

The fixed effects estimators, ϕ_j and λ_1, using the within transformation, can be written generally as:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{Y}_i \right) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}' \ddot{x}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}' \ddot{y}_{it} \right) \quad (78)$$

The health event indicator variables included in H_{it} are classified as obesity and non-obesity related in the first round of estimations, and then the shocks are included separately in the second round, as was done with the Bayesian model. The H_{it} matrix contains fewer variables than in the Bayesian model, as discussed earlier.

c. Results

Estimates of equation (77) are displayed in Table 15[84]. The grouped results for the obese show that only non-obesity related health events influence BMI. These health shocks reduce BMI by 0.7269 units. The mean BMI level for this group is 29.6, suggesting that these health experiences reduce BMI by 2.5 percent. Once the health events are separated, only congestive heart failure and stroke are significant. Since these are both obesity related events, the group must drive the significance of the grouped non-obesity health events in the first set of estimations, and not the separate conditions. Conversely, the lack of significance of the grouped obesity events, must be due to the insignificant events muting the effects of stroke and congestive heart failure. Heart failure reduced BMI by 0.9270 units, or 3.2 percent. Stroke was significant for those respondents experiencing mobility problems post health event. Thus, it is not surprising that this condition increases BMI by 1.985 units (6.3%), as the individuals report restricted movement.

In both sets of estimates, the lag of BMI, exercise, smoking and drinking status coefficients are significant. The values of the coefficients vary little between estimation results, thus the estimates from the separate health shock model are discussed here. The respondents report a negative impact of their prior BMI report. BMI in the previous interview leads to a reduction in the current report by 0.1765 units. Although the respondents use the prior when forming the posterior BMI report, its relative weight is quite small. Exercise lowers BMI by 0.3408 units, while current smokers report a 1.075

[84] The behavioral model was also estimated via POLS and POLS with a double lag, but the results are not considered here, as the level of serial correlation could not be corrected for.

unit reduction in BMI. Neither one of these results is surprising due to the increase in metabolism that occurs with the participation of each of these activities. Each drink consumed per day increases BMI by 0.1006 units. Alcohol is associated with weight gain in the literature due to its high caloric content, making this result also unsurprising.

The purpose of the behavioral model is to determine if the updated longevity perceptions from the Bayesian model are manifested in a behavioral response captured by the BMI metric. The non-obesity related health event indicator is significant in both the Bayesian and behavioral models. In the Bayesian model, the non-obesity related health event reduces the subjective report by 0.0878. At the mean, this is a 15.3 percent reduction. The added pessimism resulting from the adverse health experience leads to a reduction in BMI by 0.7269 points, or 2.5 percent. There appears, then to be a 6 to 1 belief to behavioral response ratio for the non-obesity related health events. The non-obesity related events are lung disease and cancer, neither for which weight loss is reported to improve outcome. Thus this relationship is likely due to depression and wasting from the health condition, not a behavioral response to improve outcomes. No other pairs are significant in both the Bayesian and behavioral model, thus no other comparisons can be made.

Table 15: Fixed Effects, IV Estimates, Behavioral Model		
	Normal	**Obese**
Bayesian		
BMI_lag	-0.0612	-0.1604
X		
education	0.0026	0.0019
assets	0.0515	-0.0198
male	-0.0317	0.0651
black	-0.1632	0.2298
hispanic	0.0338	0.1400
other	-0.2074	-0.0696
work	0.0628	-0.0187
unemployed	0.1245	0.1835
married	0.1301	-1.049
northeast	0.2165	0.0520
midwest	0.2005	0.1874
south	0.1800	0.1692
H		
exercise	0.0441	-0.3538
smoke	-0.4977	-1.056
drink	-0.0928	0.1054
mobility	0.1214	0.0078
Events		
heart angina	0.3634	0.2629
heart attack	-0.2143	-0.0876
conjestive heart failure	1.784	-0.9270
diabetes, insulin	5.621	0.1110
diabetes, medication	-1.611	-0.2711
high blood pressure	0.1426	-0.2673
stroke, medication	0.2717	-0.8220
stroke, problem	-0.4744	1.945
Statistics		
N	1174	2638
T	2	2
Model F-test	2.65	3.47
Prob>F	0.0000	0.0000
R-square	0.0168	0.3656
# obs.	1170	2626
corr(u_i, Xb)	-0.3726	-0.7323
Sigma_u	2.146	4.749
Sigma_e	0.7741	1.313
rho	0.8849	0.929
F-test all u_i=0	1.52	3.37
Prob>F	0.0000	0.0000

IX. Conclusion

a. Strength of Hypotheses

The Bayesian model predicts that individuals use the entire information set when forming their posterior subjective longevity probability. The information set theoretically includes a prior longevity report and any new information, including new health events. The results in this dissertation indicate that in fact neither the obese, nor the normal groups, use their prior beliefs when forming a posterior probability report[85]. Both groups do use various types of health information. The obese respond to both obesity and non-obesity related adverse health events, while the normal sample only reacts to obesity-related adverse health events. This finding does not necessarily mean that individuals are short sighted; rather that new health information is more influential than prior beliefs when considering personal survival.

The results of the Bayesian model indicate that the obese and normal groups respond differently to adverse health events. The next step of the analysis attempts to answer why these differences exist. Tests suggest that the obese assign a higher risk assessment to the obesity related health events than the normal sample does. The obese, thus, respond to a greater degree to obesity related events than the normal group. Further tests show that the obese assign a higher risk equivalent to new information than their normal peers would assign to their own experience. These results support the idea that the obese, although they choose to participate in risky behavior, do interpret new health information differently than the normal sample. To be more specific, the obese react more strongly that the normal group to new adverse health information.

[85] Given the selection of the FE-IV estimates.

The Bayesian model used in this dissertation is based upon that of Smith et al. in their attempt to determine how smokers use new health information compared to non-smokers. They found that smokers responded only to smoking related health events, although to a greater degree than non-smokers. Non-smokers responded to both smoking and non-smoking related events. A parallel finding here would be that the obese respond only to obesity related events, and more severely so than the normal group, while the non-obese react to both obesity and general health shocks. This is not the case. The driving disease in the Smith et al. work was lung disease, a smoking related event. Lung disease is a general health event in this dissertation. Lung disease is a common condition for individuals between the ages of 51 and 61. Respondents may be responding to particular diseases that they perceive to be particularly dangerous, regardless of its link to their risky behavior. The grouping of diseases as obesity and smoking related may be arbitrary in terms of the Bayesian response of the individual, and respondent's actual perception of the disease may be driving the model instead.

The behavioral model is used to determine if new health information elicits a behavioral response in addition to the Bayesian response. The behavioral response is captured through changes in BMI. The results indicate that the obese alter their weight after some obesity and non-obesity related health diagnoses. Most behavioral responses resulted in weight loss, although respondents gained weight after reporting post-stroke complications. Beliefs are the foundation of all actions. This model and its results serve as an informative extension of the Bayesian model. The Bayesian model examines the evolution of individual longevity perceptions over time, and the results show that new health information determines the direction of those beliefs. The behavioral model

establishes if this new information manifests itself in a behavioral response. The new information is hypothesized to increase the costs of obesity, and the consumption of preventative measures, including weight loss, is likely to increase. This hypothesis is supported for some health events.

b. Future Work and Refinements

The behavioral and Bayesian models are new and insightful methods of examining the obesity epidemic in the United States. There is certainly more to be done, including improvements and extensions of these models. Kan and Tsai suggest that due to the arbitrary nature of the cut-off points of the BMI classification system, it is more informative to consider the entire BMI distribution instead of discrete portions of it. They employed the quantile regression technique, and some variation of this method might yield greater insight to the models presented in this dissertation.

The only preventative measure tested in the behavioral model is weight loss. Table 7 shows several treatment options for each health event included in the dataset that either improve quality of life or reduce the probability of reoccurrence. For example, being overweight is the greatest risk factor for Type II diabetes. Fat makes the body's cells more resistant to insulin. Losing weight reverses this process, and the cells become more receptive to insulin. For some individuals, weight loss, often as little as ten to twenty pounds, is all that is needed to restore blood sugar to normal. Medications also exist that help restore blood sugar to appropriate levels. Insulin, sulfonylurea drugs, and biguanides are examples. The utility maximizing preventative measure choice may be a

drug rather than weight loss for some individuals. Respondents may be substituting less expensive preventative measures for weight loss, including medications. If this is the case, the behavioral model does, not capture preventative measures, which are also types of behavioral responses. A multi-level response model that captures a wider variety of preventative measures would be preferable to the model used in this dissertation.

Another limiting factor of the behavioral model is the absence of an insurance indicator. Insurance alters the choices of individuals. For example, and individual will consume more medical care when he or she has health insurance since he or she does not bear the entire cost. Additionally, individuals may not take as many precautionary measures when they have health insurance, as they do not pay the entire cost of an adverse health event that occur due to their risky behavior. Insurance companies use deductibles, co-payments, and other measures to avoid this type of behavior. Many aging Americans, including those in the HRS dataset, have health insurance. If moral hazard is present, but not included in the behavioral model, limits are placed on the accountability of the behavioral response captured by it. The addition of a health insurance indicator may improve the accuracy of the behavioral model.

X. References

Anderson, T.W. and Cheng Hsiao. (1981) "Estimation of Dynamic Models with Error Components." *Journal of the American Statistical Association,* 76(375), 598-606.

Anderson, T.W. and Cheng Hsiao. (1982) "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics*, 18, 47-82.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. (1994) "An Empirical Analysis of Cigarette Addiction." *The American Economic Review,* 84, 396-418.

Becker, Gary S. and Kevin M. Murphy. (1998) "A Theory of Rational Addiction." *The 1Journal of Political Economy*, 96, 675-700.

Becker, Gary S., and George J. Stigler. (1977) "De Gustibus Non Est Disputandum". *The American Economic Review*, 67, 76-90.

Boffetta, Paolo, and Lawrence Garfinkel. (1990) "Alcohol Drinking and Mortality among Men Enrolled in the American Cancer Society Prospective Study." *Epidemiology*, 1, 343-348.

Chaloupka, Frank. (1991) "Rational Active Behavior and Cigarette Smoking." *The Journal of Political Economy*, 99, 722-742.

Chou, Shin-Yi, Michael Grossman, and Henry Saffer. (2002) "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." *National Bureau of Economic Research Working Paper*, 9247.

Cutler, David M., Edward L. Glaeser, and Jesse M. Shapiro. (2003) "Why Have Americans Become More Obese?" *National Bureau of Economic Research Working Paper*.

Denes-Raj, Veronika, Howard Ehrlichman. (1991) "Effects of Premature Death on Subjective Life Expectancy, Death Anxiety, and Health Behavior." *The Journal of Death and Dying*, 23(4), 309-321.

Diamond, Peter and Botond Koszegi. "Hyperbolic Discounting and Retirement." Mimeo, 1998.

Dow, William H., Tomas Philipson, and Xavier Sala-i-Martin. (1999) "Longevity Complementarities Under Competing Risks." *American Economic Review,* 89(5), 1358-1371.

Ellison, R. Curtis. (1990). "Cheers!" *Epidemiology,* 1, 337-339.

Evans, William N., W. Kip Viscusi, "Estimation of State-Dependent Utility Functions Using Survey Data," *The Review of Economics and Statistics* 73 (February 1991), 94-104.

Feinstein, J. (1993) "The Relationship Between Socioeconomic Status and Health: A Review of the Literature." *The Milbank Quarterly,* 71, 279-322.

Flegal, Katherine M., PhD., Margaret D. Carroll, M.S., Cynthia L. Ogden, PhD, Clifford L. Johnson, MSPH. (2002). "Prevalence and Trends in Obesity Among US Adults, 1999-2000." *Journal of the American Medical Association*, 288, 1723-1727.

Greene, William H. 2003. *Econometric Analysis*. New Jersey: Pearson Education, Inc.

Grossman, Michael. 1972. *The Demand for Health*: *A Theoretical and Empirical Investigation*. New York: Columbia University Press.

Gruber, Jonathan and Botond Koszegi. (2000) "Is Addiction 'Rational'? Theory and Evidence." NBER Working Paper 7507.

Gujarati, Damodar N. (1995) *Basic Econometrics.* New York: McGraw-Hill, Inc.

Hamermesh, Daniel S. (1985) "Expectations, Life Expectancy, and Economic Behavior," *Quarterly Journal of Economics*, 100, 389-408.

Harris, Christopher and David Laibson. (2001) "Dynamic Choices of Hyperbolic Discounters." *Econometrica*, 69, 935-957.

Hausman, J. A. (1978). "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

Hurd, Michael D., Kathleen McGarry. (1995) "Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study," *The Journal of Human Resources.* Volume 30, Special Issue on the Health and Retirement Study: Data Quality and Early Results, S268-S292.

Hurd, Michael D., Kathleen McGarry. (2002) "The Predictive Validity of Subjective Probabilities of Survival," *Economic Journal.* Volume 112, 966-985.

Juster, F. Thomas, Richard Suzman. (1995) "An Overview of the Health and Retirement Study." *The Journal of Human Resources*, 30, S7-S56.

Kan, Kamhon, Wei-Der Tsai. (2004) "Obesity and Risk Knowledge." *Journal of Health Economics*, 23, 907-934.

Kitagawa, Evelyn M., and Philip M. Hauser. (1973) *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology.* Cambridge, MA: Harvard University Press.

Laibson, David. (1996) "Hyperbolic Discount Functions, Undersaving, and Savings Policy." *National Bureau of Economic Research Working Paper*.

Laibson, David. (1997) "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 62, 443-477.

Lakdawalla, Darius and Tomas Philipson. (2002) "The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination." *National Bureau of Economic Research Working Paper*, 8946.

Lichtenstein, Sarah, Paul Slovic, Baruch Fishoff, Mark Layman, and Barbara Combs. (1978). "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory*, 4, 551-579.

Marshall, Alfred. "Principles of Economics". 8th edition, London 1923, p.94.

Mundlak, Y. (1978). "On the Pooling of Time Series and Cross Sectional Data," *Econometrica*, 46, 69-85.

NIH, NHLBI. (1998) "Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults." HHS, PHS, 23.

Peltzman, Sam. (2001) "Offsetting Behavior and Medical Breakthroughs," George J. Stigler Center for the Study of the Economy and the State, University of Chicago, Working Paper No. 169.

Philipson, Tomas J., and Richard A. Posner. (1999) "The Long Run Growth in Obesity as a Function of Technological Change." *National Bureau of Economic Research Working Paper* 7423.

Picone, Gabriel, Frank Sloan, and Donald Taylor. (2004) "Effects of Risk and Time Preference and Expected Longevity on Demand for Medical Tests." *Journal of Risk and Uncertainty*, 28(1), 39-53.

Shaper, A.G. (1990) "Alcohol and Mortality: A Review of Prospective Studies." *British Journal of Addiction*, 85, 837-847.

Silver, Morris. (1972) "An Econometric Analysis of Spatial Variations in Mortality Rates by Age and Sex." In *Essays in the Economics in Health and Medical Care,* ed. Victor Fuchs, 161-227. New York: Columbia University Press.

Smith, V. Kerry, Donald H. Taylor, Jr., and Frank Sloan. (2001) "Longevity Expectations and Death: Can People Predict Their Own Demise?" *American Economic Review*, 91(4), 1126-1134.

Smith, V. Kerry, Donald H. Taylor, Jr., Frank A. Sloan, F. Reed Johnson, and William H. Desvousges. (2001) "Do Smokers Respond to Health Shocks?" *The Review of Economics and Statistics,* 83, 675-687.

St. Clair, Patricia, et al. "RAND HRS Data Documentation," Version B, April 2003, prepared for the Social Security Administration, pg. 2.

Strum, R. and KB Wells. (2001) "Does obesity contribute as much to morbidity as poverty or smoking?" *Public Health* 115, 229-235.

U.S. Department of Health and Human Services. "The Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity." [Rockville, MD]: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General: [2001].

Viscusi, W. Kip. (1985) "A Bayesian Perspective on Biases in Risk Perceptions," *Economics Letters* 17, 59-62.

Viscusi, W. Kip and Charles O'Connor, "Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?" *American Economic Review* 74:5 (1984), 942-956.

Viscusi, W. Kip and John Hakes. (2003). "Risk Ratings that do not Measure Probabilities," *Journal of Risk and Insurance*, 6(1), 23-43.

Viscusi, W. Kip and Michael J. Moore. (1989). "Rates of Time Preference and Valuations of the Duration of Life," *Journal of Public Economics*, 8(3), 397-417.

Viscusi, W. Kip and William N. Evans, (1998) "Estimation of Revealed Probabilities and Utility Functions for Product Safety Decisions," *The Review of Economics and Statistics,* 80, 28-33.

Wooldridge, Jeffery M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: The MIT Press.

Wooldridge, Jeffery M. 2004. *Introductory Econometrics: A Modern Approach, Third Edition*. Mason: Thomson-South-Western.

World Health Organization. Obesity: Preventing and Managing the Global Epidemic. Geneva: World Heath Organization, 1997.

XI. APPENDICES

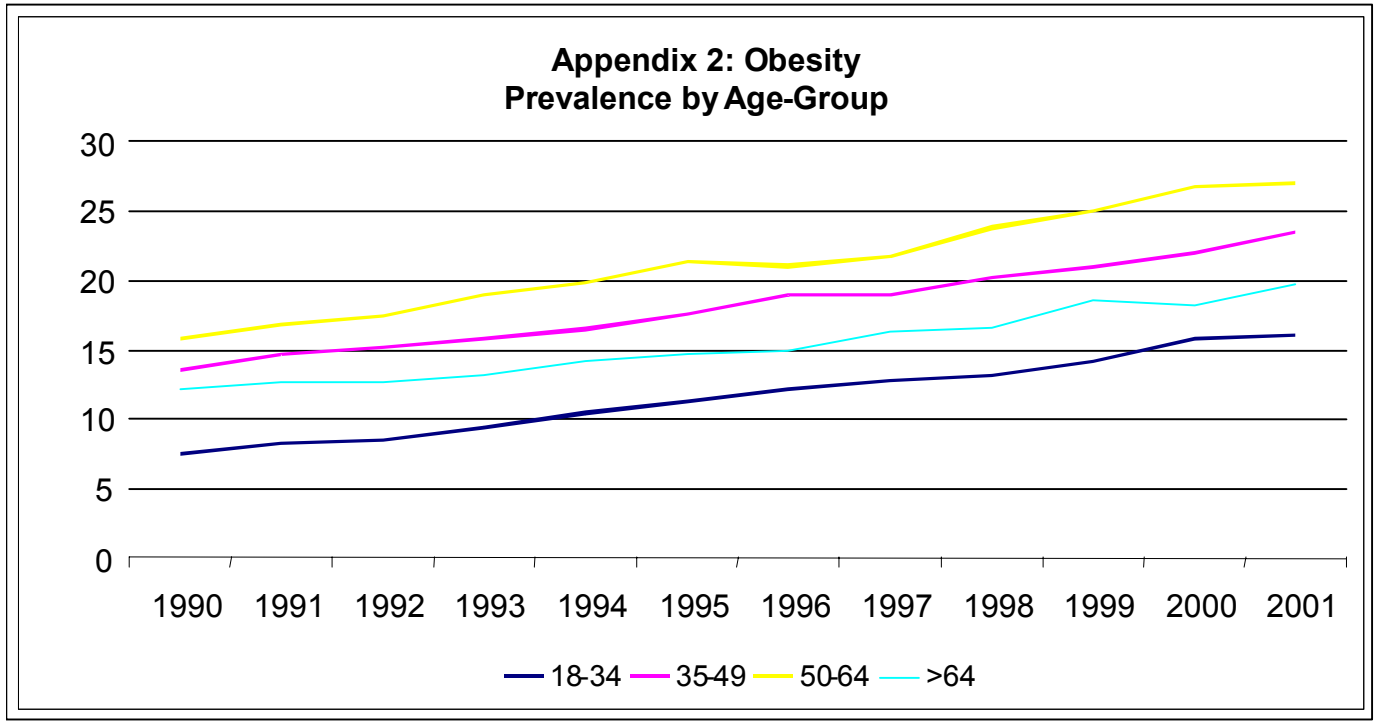
Appendix 1: Trends in the Age-Adjusted and Age Specific Prevalence of Obesity for Adults Aged 20-74 Years, 1960-2000*

		Prevalence, %					Change, % (95% CI)***	
Sex	Age, y**	NHES I, 1960-1962 (n=6126)	NHANES I, 1971-1974 (n=12911)	NHANES II, 1976-1980 (n=11765)	NHANES III, 1988-1994 (n=14468)	NHANES Continuous, 1999-2000 (n=3601)	NHANES II to NHANES III	NHANES III to NHANES 1999-2000
Both Sexes	20-74	13.4	14.5	15.0	23.3	30.9	8.3 (6.6-10.0)	7.6 (4.2-11.0)
Men	20-74	10.7	12.1	12.7	20.6	27.7	7.9 (6.0-9.8)	7.1 (3.4-10.8)
	20-39	9.8	10.2	9.8	14.9	23.7	5.1 (2.9-7.2)	8.8 (4.8-12.8)
	40-59	12.6	14.7	15.4	25.4	28.8	10.0 (6.9-13.0)	3.4 (-2.8-9.6)
	60-74	8.4	10.5	13.5	23.8	35.8	10.3 (6.3-14.3)	12.0 (5.0-19.0)
Women	20-74	15.8	16.6	17.0	25.9	34.0	8.9 (6.5-11.3)	8.1 (3.7-12.5)
	20-39	9.3	11.2	12.3	20.6	28.4	8.3 (5.2-11.4)	7.8 (2.5-13.1)
	40-59	18.5	19.7	20.4	30.4	37.8	10.0 (6.1-13.9)	7.4 (0.5-14.3)
	60-74	26.2	23.4	21.3	28.6	39.6	7.3 (3.9-10.6)	11.0 (4.6-17.4)

*NHES Indicates national Health Examination Survey; NHANES, National Health and Nutrition Examination Survey; and CI, confidence interval.

**Estimated prevalences for ages 20-74 years were age standardized by the direct method to the 2000 Census population using age groups 20-39, 40-59, and 60-74 years.

***Overall and within each age-sex group, the chances between 1988-1994 and 1999-2000 are not significantly different from the changes between 1976-1980 and 1988-1994.


Appendix 2: Obesity
Prevalence by Age-Group
30
25
20
15
10
5
0
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
18-34
35-49
50-64
>64

Appendix 3: BMI Classification Table

Height in Feet and Inches	Weight in Pounds													
	120	130	140	150	160	170	180	190	200	210	220	230	240	250
4'6	29	31	34	36	39	41	43	46	48	51	53	56	58	60
4'8	27	29	31	34	36	38	40	43	45	47	49	52	54	56
4'10	25	27	29	31	34	36	38	40	42	44	46	48	50	52
5'0	23	25	27	29	31	33	35	37	39	41	43	45	47	49
5'2	22	24	26	27	29	31	33	35	37	38	40	42	44	46
5'4	21	22	24	26	28	29	31	33	34	36	38	40	41	43
5'6	19	21	23	24	26	27	29	31	32	34	36	37	39	40
5'8	18	20	21	23	24	26	27	29	30	32	34	35	37	38
5'10	17	19	20	22	23	24	26	27	29	30	32	33	35	36
6'0	16	18	19	20	22	23	24	26	27	28	30	31	33	34
6'2	15	17	18	19	21	22	23	24	26	27	28	30	31	32
6'4	15	16	17	18	20	21	22	23	24	26	27	28	29	30
6'6	14	15	16	17	19	20	21	22	23	24	25	27	28	29
6'8	13	14	15	17	18	19	20	21	22	23	24	25	26	28

Appendix 4: Data Filters

Interaction	Variable	Outcome
(Heart angina)*(Medication Angina)	hshrtan	=1 if heart angina & medication
(Heart attack)*(Year of heart attack)	hshrtatt	=1 if heart attack & year
(CHF)*(Medication for CHF)	hshrtchf	=1 if chf & medication
(Arthritis since)*(Medication arthritis)	hsarth	=1 if arthritis & medication
(Cancer since)*(Treatment)	hscancr	=1 if cancer & treatment
(Diabetes since)*(Insulin)	hsdiabin	=1 if diabetes & insulin
(Diabetes since)*(Medication diabetes)	hsdiabmed	=1 if diabetes & medication
(HBP since)*(Medication HBP)	hshbp	=1 if hbp and medication
(Lung since)*(Medication/Treatment lung)	hslung	=1 if lung & medication/treatment
(Psych since)*(Medication psych)	hspsych	=1 if psych & medication
(Stroke since)*(Stroke medication)	hsstkm	=1 if stoke & medication
(Stroke since)*(Problems from stroke)	hsstkp	=1 if stroke & problems

Appendix 5: Response Attrition

Dataset	# Obs.	Loss	% of Previous	% of Total
HRS-Rand	10167	___	___	___
Healthy	7358	2809	0.7237	0.7237
Obese	4700		0.6388	0.4623
Normal	2440		0.3316	0.2400
Missing Obs.	1907	5451	0.7408	0.1876
Obese	1319	3381	0.2806	0.1297
Normal	588	1852	0.2410	0.0578

Appendix 6: Risk Factors for Adverse Health Events							
	Congestive Heart Failure	**Stroke**	**Diabetes Mellitus**	**High Blood Pressure**	**Heart Attack**	**Lung Disease**	**Arthritis**
Unavoidable Risk Factors							
Age	x	x	x	x	x	x	x
Gender	x	x	x		x		x
Heredity	x	x	x	x	x	x	x
Prior Stroke	x	x					
Prior Heart Attack	x	x			x		
Avoidable/Treatable Risk Factors							
Smoking	x	x		x		x	x
High Cholesterol	x	x	x		x		
HBP	x	x	x	___	x		
Physical Inactivity	x	x	x	x	x		
Overweight/Obesity	x	x	x	x	x		x
Blood Disorders		x					
Heart Disease	x	x			x		
Diabetes	x	x	___		x		
Other							
Alcohol	x	x			x		
Stress	x	x			x		

Appendix 7: Treatment Options for Health Events							
	Congestive Heart Failure	Stroke	Diabetes Mellitus	High Blood Pressure	Heart Attack	Lung Disease	Arthritis
Lifestyle Changes							
Diet	x		x	x	x		
Exercise	x		x	x	x		x
Weight Loss	x	x	x	x	x		x
Alcohol Consumption	?	moderate	?	?	?		
Smoking Cessation	x	x		x	x	x	
Control of Comorbidity	x						
HBP	x	x		___	x		
High Cholesterol	x	?			x		
Heart Disease	___	x	x	x	x		
Diabetes	x	?	___		x		
Medications	x	x	x	x	x	x	x
Physical Therapy/Rehabilitation						x	x
Transplantation/Surgery	x	x	x		x	x	