NBER WORKING PAPER SERIES

SELF-REGULATION AND HEALTH

Henry Saffer

Working Paper 20483
http://www.nber.org/papers/w20483

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2014

Excellent research assistance was provided by Markus Gehrsitz. Michael Grossman provided many helpful comments. The work on this project has been funded by grant R01 AA020464 from the National Institute on Alcohol Abuse and Alcoholism, NIH to the National Bureau of Economic Research. This grant is supported by the Common Fund, which is managed by the OD/Office of Strategic Coordination (OSC). The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Self-regulation and Health
Henry Saffer
NBER Working Paper No. 20483
September 2014
JEL No. I1

ABSTRACT

The purposes of this paper are to measure self-regulation, to investigate whether self-regulation differs across different health related choices, to estimate its effect on health choices and to estimate the effect of self-regulation on health-demographic gradients. The theory and empirical approach to self-regulation employed in this paper relies on a broad literature which includes economics, psychology and experimental studies. In addition, a novel empirical approach is employed to create a single measure of self-regulation that can vary across domains. A single measure of self-regulation in place of a set of proxy variables allows for the study of how self-regulation is correlated across different health choices. The results show that there is a high correlation in self-regulation for smoking, drinking, drug use, crime and gambling, but that self-regulation for BMI (body mass index) and obesity are different than self-regulation for the other outcomes. The results show that self-regulation has a significant negative effect on all choices. The results also show that self-regulation generally reduces the effect of education on health but education retains a negative and significant relationship with all outcomes. The research presented in this paper also raises questions about the effect of omitted individual heterogeneity in measuring the effects of public policy.

Henry Saffer
NBER
5 Hanover Square, 16th Floor
Suite 1602
New York, NY 10004-2630
hsaffer@gc.cuny.edu

1. Introduction

Recent studies by economists have examined what is known as noncognitive skills and their effect on health. These noncognitive skills are also referred to as time preference, personal efficacy, self-esteem, locus of control, organizational skills, self-regulation, motivation, adventurousness, self-control, conscientiousness and socioemotional skills (Chiteji, 2010).[1] One concern with these past studies is that they often take an ad hoc approach to noncognitive skills.[2] There is, however, a literature outside of economics which examines the effects of noncognitive skills on health related outcomes and provides a theoretical and empirical approach to the measurement of noncognitive skills. In this paper the noncognitive skill of interest is the individual's ability to defer an immediate reward for a future reward which will be referred to as self-regulation (SR). Because an individual can exhibit a preference for a future outcome (time preference) only to the degree that they can resist the desire for immediate gratification (self-regulate) the concepts of time preference and SR are closely aligned.

One advantage in studying SR rather than time preference is that it bypasses the assumptions of time preference theory. Time preference theory is described by the discounted utility (DU) model. The assumptions in the DU model are analytically convenient, but according to Frederick, et al. (2002), are not consistent with empirical observation. A central assumption of the DU model is that all of the motives underlying time preference can be represented by a single discount rate and that utility from all sources is discounted at the same rate. Studies have found that discount rates vary by outcome, gains are discounted more than losses, small amounts discounted more than large amounts, and explicit sequences of multiple outcomes are discounted differently than outcomes considered individually. Another concern with the DU

[1] The term non-cognitive skill is somewhat inaccurate in that the processes described involve a balance of reasoning and emotion.
[2] A notable exception to this is Conti and Heckman (2010).

model is that observed discount rates appear to decline in future time periods. Discount rates can be modeled with a hyperbolic discount function to account for this phenomenon. While this approach provides an important mathematical restatement, it does not provide a theoretical explanation of the phenomena. Frederick, et al. (2002) also report that the literature provides no consensus for quantitative estimates of discount rates. These concerns generally do not relate to SR. More detail on the theory of SR is provided in section 3.

The purpose of this paper is to measure SR, to investigate whether it differs across different health choices and to estimate its effect on health choices. The empirical work also includes crime and gambling as outcomes because these outcomes are also thought to be related to SR and provide additional empirical tests of the theory. The outcomes are commonly referred to as domains. The theory and empirical approach to SR relies on a broad literature including economics, psychology and experimental studies. In addition, a novel empirical approach is employed to create a single latent variable measure of SR. This single measure of SR is not restricted to the same value for all outcomes as is the case in all prior studies. A single outcome specific measure of SR allows for the study of how SR is correlated across different health choices. This study of SR is important because SR and public policy are alternative mechanisms which can affect health. Public policy becomes more important for choices in which SR is low. Also, public policies and incentives which improve health behavior in one domain can enhance SR and may affect SR in closely related domains.

2. Prior studies

There are several recent empirical studies in the economics literature that measure the effect of noncognitive skills on health domains. The emphasis in these studies is on how noncognitive skills moderate the health-education gradient. A recent and extensive study of the health-education gradient was provided by Cutler and Lleras-Muney (2010). They use data from several surveys from two countries and find that income, health insurance, and family background can account for about 30 percent of the gradient. Knowledge and measures of cognitive ability explain an additional 30 percent. Social networks account for another 10 percent. Proxies for time discounting and risk aversion do not account for any of the education gradient, and neither do personality factors such as a sense of control of oneself or over one's life. A set of 16 questions on future health and financial domains are employed to measure time discounting.

Conti and Hansman (2013) reexamine the personality traits studied by Cutler and Lleras-Muney (2010) and add the Rutter Behavior Scale and a number of syndrome scores from the British Social Adjustment Guide (BSAG). The Rutter Scale measures behavioral difficulties and was constructed by summing the responses to a series of questionnaire items, with a higher score indicating behavioral problems. The BSAG was designed to describe a child's behavior in particular settings. The questionnaire results in 12 different measures of social maladjustment. Conti and Hansman (2013) conclude that personality contributes to the education-health gradient to an extent nearly as large as that of cognition.

Another study by Chiteji (2010) employs data from the Panel Study of Income Dynamics (PSID) to ascertain the relationship between a number of different socio-emotional attributes and drinking and exercising. The socio-emotional attributes analyzed are the degree to which an individual is future-oriented, and self-efficacy. The measure of future-orientation comes from a question that the PSID asked about the length of the individual’s time horizon, and self-efficacy is an index of several questions relating to the evaluation of one’s ability to be effective performing tasks that are necessary to realize an outcome. Chiteji (2010) finds that future-orientation and self-efficacy have a negative effect on drinking and a positive effect on exercise.

Conti, Heckman and Urzua (2010) estimate the effects of cognitive, noncognitive, and early health endowments on later health outcomes using data from the British Cohort Study.

They find that not accounting for noncognitive ability results in overestimation of the importance of cognitive ability in determining later health. They use six scales as measurements of noncognitive ability: the child was asked the locus of control questions and five scales were administered to the teacher (perseverance, cooperativeness, completeness, attentiveness, and persistence). Selection on preexisting traits explains more than half of the observed differences in poor health and obesity. They also find that not accounting for noncognitive ability overestimates the importance of cognitive ability in determining later health.

Kaestner and Callison, (2011) examine the effects of noncognitive traits measured at the end of childhood on mental and physical health at age 41 using the National Longitudinal Survey of Youth 1979. They find that self-esteem has a significant association with health at age 41 and some evidence that the internal locus of control is associated with better adult health. The internal locus of control refers to the extent to which individuals believe that their choices can affect outcomes in their life. They also conclude that differences in the noncognitive factors they studied are not important explanations of gender or racial differences in health.

Savelyev (2014) investigates how education, cognitive skills and personality skills affects health. The data employed are the 1922–91 Terman lifecycle sample of children with high ability. Using factor analysis to create personality variables he estimates a model that uses cognition, personality, and education as arguments of the production function for health. The model also includes interaction between education and personality in producing health. Gender specific models are estimated. For males, there are strong effects of personality, IQ, education and the interaction of personality and education. For females the effects of education and personality are not precisely determined.

The literature outside of economics provides a number of empirical studies which have investigated the effect of self-regulation proxies on negative health behaviors.[3] Moffitt et al. (2011) used data which followed a cohort of 1,000 children from birth to the age of 32 and found that low self-regulation in childhood predicts low physical health, substance dependence, problems with personal finances, and criminal behavior. Effects of the individual's self-regulation were differentiated from intelligence, social class and mistakes made in adolescence. Moffitt et al. (2011) also investigated a cohort of 500 sibling-pairs. The sibling with lower self-regulation had poorer outcomes, despite shared family background. A study by MacKillop and Kahler (2009) examined the effect of low self-regulation on smoking behavior. They define low self-regulation as an index of impulsivity that reflects a preference for smaller immediate rewards over larger delayed rewards (i.e. a present orientation). Current smokers exhibit lower self-regulation compared to nonsmokers but also exhibit lower self-regulation compared to ex-smokers, suggesting that either self-regulation is positively associated with successful smoking cessation or that smoking cessation itself increases self-regulation. Gubler and Pierce (2014) examine the relationship between contributions to a 401(k) retirement plan and individual health choices. Individuals in the study were provided with a physical health examination and some discovered that they had abnormal blood test results. This health examination is a quasi-exogenous information shock. The authors controlled for initial health, demographics, job type, and income and found that individuals who contributed to a 401(k) plan were 27% more likely to change their behavior to improve their health than non-contributors. These findings are consistent with an underlying individual time discounting trait that is both difficult to change, similar across domains and distinct from the trait known as conscientiousness.

[3] Studies of self-regulation outside of economics include: Gailliot and Baumeister, 2007; Wright et al., 2012; Belsky and Beaver, 2011; Beaver, Ratchford and Ferguson, 2009; Stanovich and West, 2008; Vaughn et al., 2009; Miller, Barnes and Beaver, 2011; Jensen-Campbell et al., 2007. These studies are discussed in more detail in section 3.

An important issue highlighted in prior studies is the potential variation in SR across different domains. That is, for a given individual, to what degree is SR the same across domains? This issue is in contrast to the assumption in the DU model that utility from all sources is discounted at the same rate. If SR differs significantly across domains then a SR measure derived from one domain such as financial choice may not be relevant to other domains such as health. Several prior studies have investigated the domain specificity of self-regulation. Foxall et al. (2011) studied consumers making hypothetical decisions with respect to financial returns, health domains and vacation alternatives. They find that self-regulation differs across these broadly defined domains. Jimura et al. (2011) examined the relation between discounting of hypothetical money and real liquid rewards in young adults and older adults. At the individual level, the rates at which young and older participants discounted each reward type were stable over a two to fifteen week interval. However, they find no correlation in self-regulation across these broadly defined domains. These results suggest that, although similar decision-making processes may underlie discounting in different domains, the rates at which individuals discount money and consumable rewards are a domain specific process. Odum (2011) reviews a number of studies and also finds that money is generally discounted less than drugs, food, alcohol and cigarettes and that self-regulation is relatively stable over time, while exhibiting developmental and experiential change. Odum (2011) also finds that self-regulation is associated with activity in particular brain regions, certain genetic traits, and appears to be heritable. Odum (2011) concludes that self-regulation is a stable and pervasive individual characteristic and that it is strongly correlated across narrowly defined domains.

Another related issue is the stability of SR for a given individual. Casey, Galvan and Hare (2005) find that impulsivity peaks at about age 16 and that most individuals achieve adult levels of self-regulation in their mid-20s. However, prior studies have suggested that interventions can create long term changes in SR and that changes in SR in one health area can also spillover to other health areas. Deckersbach et al. (2014) found that individuals in a weight loss program for six months had a shift in brain activation on an MRI test in favor of low calorie relative to high calorie foods. This study suggests that following a healthy diet may, in time, result in an increased preference for healthy food. Charness and Gneezy (2009) study the effect of financial incentives to attend a gym on future outcomes. In two experiments they paid individuals to attend a gym. After the intervention ended they found that gym attendance increased and also found improvements in health indicators. This study suggests that a period of incentivized increase in SR can lead to an actual increase in SR. Baumeister et al., (2006) review the research on energy intensive aspect of SR. Various activities including SR deplete this energy making added SR more difficult. This is referred to as ego depletion. However, they argue that if an individual chooses to exercise greater self-regulation or is incentivized by some external factors then this can produce improvements in the ability to self-regulate similar to strengthening a muscle. These findings are consistent with Becker and Mulligan's (1997) theory of endogenous time preference.

3. Estimation

The empirical model is based on a demand equation which is derived by assuming that a consumer maximizes a utility function subject to a budget constraint. Health and other goods are arguments of the utility function. Income and full prices are represented by the income constraint. Work by psychologists, behavioral economists and neuroeconomists provide a theoretical basis for including SR in the decision process. SR is assumed to be an energy intensive neurological activity which involves the conscious suppression of the desire for immediate gratification (Fudenberg and Levine, 2006). Assume that the full price of health includes the money price and the price of self-regulation. Individuals who are able to self-regulate at a lower cost will have a lower full price for health. The maximization problem results in an equation which shows that the demand for health is a function of its money price, income and self-regulation. SR enters the demand for health with an expected positive sign. The demand for a negative health good such as cigarettes is then derived from the demand for health. The demand for this good is a function of its price, self-regulation skill, income and other variables. Aggregating across consumers results in the market demand function. The demand for the good C is specified as a function of prices, education, self-regulation, income and demographic variables. Prices and public policies are controlled with year and state fixed effects variables, ts. ED is defined as education, SR is defined as self-regulation, income and demographic variables are included in Z, and ε is a random error term. The empirical equation is:

$$C_i = \pi_0 + \pi_1 ts + \pi_2 ED_i + \pi_3 SR_i + \pi_4 Z_i + \varepsilon_i. \quad (1)$$

Equation (1) cannot be directly estimated because SR is a latent variable. One approach when a latent variable is included in a regression is to include a set of proxy variables in place of the latent variable. However, a single SR construct, is preferred to a set of proxies because it is easier to interpret one coefficient than a set of coefficients of potentially collinear variables. Two older approaches to creating a single construct are first principal component analysis and factor analysis. These approaches extract a common component of all the proxy variables to create a composite variable. However, these extracted components are redundant and generally do not maximize the predictive power of the proxy variable set. Another approach to creating a single construct is to add or average the proxy variables, possibly standardized by their standard deviations. Proxy variables measure the latent variable but with an unknown multiplicative and additive error structure. This error results in attenuation bias. When the proxy variables are simply added together, more variables increase the potential information but also increase the total error. For this approach to extract the largest possible information requires the researcher to know the relative degree of measurement error in each proxy variable, as well as the correlation in measurement error across the proxies. Unfortunately these measures are not known.

An alternative approach to constructing a single measure of a latent variable from a set of proxies and which does not require knowledge of the measurement error is provided by Lubotsky and Wittenberg (2006). They derive a set of weights which minimizes the measurement error bias and which can be used to both estimate the coefficient of SR, π_3, and to construct a single measure of SR. This estimator is referred to as the post-hoc estimator (PH). There are i individuals and j proxy variables and x_{ji} is a proxy variable for self-regulation. The PH estimator of π_3 is denoted as π^{PH} and requires the estimation of:

$$C_i = \pi_0 + \pi_1 ts + \pi_2 ED_i + \pi_3 Z_i + \sum b_j x_{ji} + \varepsilon_i. \quad (2)$$

π^{PH} is computed as: $\pi^{PH} = \sum [cov(y,x_{ji}) / cov(y,x_{1i})] b_j$. (3)

The choice of which variable to use as proxy 1 is arbitrary but provides the metric for the latent variable. The PH estimator is a lower bound rather than a point estimate for the coefficient of the latent variable. The single summary measure of SR (SR^{PH}_i) is calculated as:

$$SR^{PH}_i = \sum (b_j x_{ji}) / \pi^{PH} \quad (4)$$

SR^{PH}_i replaces SR_i in equation (1) which results in:

$$C_i = \pi_0 + \pi_1 ts + \pi_2 ED_i + \pi_3 Z_i + \pi^{PH} SR^{PH}_i + \varepsilon_i. \quad (5)$$

Because the latent variable, SR^{PH}, is simply a weighted average of the proxy variables, the coefficients π_0, π_1, π_2 and π_3 are the exactly the same in equation 2 as in equation 5. The weights are based on the marginal effects of the proxies on the outcome, holding all the other proxies and independent variables constant. The marginal effects of the proxies differ across each domain which results in a different estimated SR variable for each domain. The relationship between the SR latent variables for each domain is examined in section 5.

4. Data

This study uses the National Longitudinal Study of Adolescent Health (Add Health) data to estimate the empirical model. The Add Health is a longitudinal, nationally representative sample of individuals who were adolescents, in grades 7-12, in the United States, during the 1994-95 school year (wave 1). The in-home component included interviews with 20,745 adolescents. In 1996, (wave 2), 14,738 individuals were reinterviewed. In 2001-2002, (wave 3), 15,159 of the original wave 1 respondents were reinterviewed. In 2007-2008, (wave 4), 15,701 of the original wave 1 sample were reinterviewed. The working data set was created with data from wave 3 and wave 4, with the exception of maternal attachment which ideally should be measured at younger ages. The final working data set includes two observations on about 13,000 individuals. Add Health includes health domains such as tobacco and alcohol consumption, demographic data, income and a number of potential proxies for SR. There are 11 domains in the areas of: 1) smoking, 2) alcohol consumption 3) illicit drug use 4) caloric consumption 5) criminal behavior and 6) gambling. Minorities are oversampled and weights are included in the data set to create correct sample means. All of the estimation models include state clustered Huber standard errors based on pseudo-state variables that allow clustering by state, although the state is not identified.

4.1 Domains

The Add Health has a number of questions relating to tobacco use and alcohol use. A tobacco use variable was constructed from two questions asked in each wave. The first question is “During the past 30 days, on how many days did you smoke cigarettes?” This question has 0-30 as response codes. The second question is “During the past 30 days, on the days you smoked, how many cigarettes did you smoke each day?” This question has 1-100 as response codes. These variables were multiplied together to form an estimate of monthly tobacco consumption. Similarly, an alcohol use variable was constructed from two questions asked in each wave. The first question is “During the past 30 days, on how many days did you drink?' There are seven response codes from none to every day or almost every day. This was recoded into a continuous variable from 0 to 30. The second question asks about the number of drinks usually consumed each time on each drinking occasion in the past 30 days. The response codes are from 1-18. These variables were multiplied to form an estimate of monthly alcohol consumption.[4] The Add Health also has a measure of binge drinking which is defined as five or more drinks for a male and four or more drinks for a female. A dichotomous binge drinking variable was defined as equal to one for individuals who reported that they engaged in binge drinking 2 or more times per month.

The Add Health also has questions relating to illicit drug use. Two illicit drug use variables were defined. The first is the number of times the individual used marijuana in past 30 days. The second is the number of drug abuse symptoms which ranges from zero to four.

The Add Health also has data on height and weight which can be used to compute BMI. These data are measured by Add Health which eliminates error due to self-reporting bias. In wave 3, height is reported in feet and inches and weight is reported in pounds. The formula used to compute BMI for these three waves is weight (in pounds) / height^2 (in inches) all times 703. BMI is directly reported in wave 4. An additional dichotomous variable, obesity, is defined for a BMI equal to 30 or more.

The Add Health also has questions relating to non-health behaviors relating to self-regulation. These domains are criminal behavior and gambling. Two criminal behavior variables were coded. The first is a dichotomous construct equal to one if the individual reports that they have ever been arrested by the police. The second measures participation in various criminal activities including burglarizing, minor theft, major theft, use of a weapon, selling drugs,

[4] A dichotomous drinking variable was not included because the majority of adults drink with no negative health consequences and those who do not drink have been found, in some studies, to be less healthy than drinkers.

gang fighting, property damage, causing serious injury, pulling a weapon on someone, stabbing/shooting someone. Each of these questions were coded as dichotomous variables and then they were added together to form a crime index which goes from zero to 10. The gambling variable was defined as equal to one if the individual reported that they had ever had a net loss from gambling of more than $500 in one year.

4.2 Proxy Variables

The SR^{PH}_i variable created by (4) is an index of the effects of the proxy variables on the domain variable. The choice of proxy variables was based on a theoretical and empirical literature on SR which supports the assertion that these measures are proxies for self-regulation. As noted above, SR is an energy intensive process of suppressing the desire for immediate gratification in favor of a future reward. Cunha and Heckman (2008) employ a production function to describe this process. Ideally, the proxy variables should be measures of the production determinants which include the efficiency of the production process and the inputs in the production process. The specific proxies were chosen based on prior studies which made a theoretical argument or provided empirical evidence that the variable affected a domain related to self-regulation. These studies are discussed in more detail below. Equation (4) can be thought of as a production function where output in not observed but is estimated as a linear transformation of the production determinants and correlates of production. The index is labeled SR^{PH}_i because the evidence suggests that the proxies affect an individual's ability to delay gratification although there is no way to definitively prove that SR^{PH}_i actually measures self-regulation.

Unlike prior studies of the health gradient by economists, this paper takes a more systematic approach to the choice of proxy variables. The proxy variables chosen are measures of the efficiency and inputs in the process of producing self-regulation. The efficiency of production process is related to the individual's biological endowment and other variables.

Any factor which alters production efficiency can affect the production of self-regulation. The primary input in SR is energy which is a limited resource.[5] Other processes requiring energy can take priority over self-regulation reducing the energy available for self-regulation.[6] Thus proxy measures for competing processes which require energy also belong in equation (4). The proxy variables and the evidence provided by prior studies which support the assertion that these measures are proxies for the determinants of SR are discussed next. Some of these empirical studies also use the Add Health data. More detailed information on the empirical definitions of the proxy variables is provided in the Appendix 1.

Wave 3 and wave 4 of the Add Health each provide good proxy variables for self-regulation.[7] All of these proxy variables have been shown in prior studies to affect self-regulation. An important question is whether SR between these two waves can be assumed to be constant. Past studies have concluded that individuals reach a stable adult levels of SR at about age 25. Moffitt et al. (2011) measured self-regulation with data from five separate waves of the Dunedin Study sample. They also compared self-regulation measured in childhood with a measure derived more than 10 years later and found that the rank order in self-regulation did not change. In the Dunedin sample, the measures of self-regulation from childhood predicted health domains, wealth domains and criminal conviction history at age 32. These results also parallel the results found by Mischel, Shoda and Peake (1988). Odum (2011) reviews a number of studies and finds that self-regulation is associated with activity in particular brain regions, certain genetic traits, and appears to be heritable. She concludes that self-regulation is a stable and pervasive individual characteristic which can exhibit developmental and experiential

[5] Gailliot and Baumeister (2008) argue the energy is glucose.
[6] For example, any process required for survival.
[7] One exception is a retrospective parental variable from wave 1 which asks the individual to recall earlier childhood experiences with their mother.

change. Developmental and experiential change will, to some degree, be controlled with observable demographics such as marital status and parent status in the regressions. According to Heckman (2008) non-cognitive traits such as self-regulation are largely determined at young ages and do not change significantly after adolescence. The individuals in the Add Health are on average about age 23 in wave 3 and about age 29 in wave 4. Based on the evidence, it seems reasonable to assume that SR is sufficiently constant between waves 3 and 4 that a time invariant index of SR derived from wave 3 and wave 4 data will not introduce any important bias.

The Add Health has a question set designed to measure self-regulation independent of any specific domain. The questions focus on impulsivity, feelings of decreased regulation and instinctive rather than thoughtful decisions. This variable will be called the self-regulation proxy and the results of equation (4) will be called the SR latent variable. The self-regulation proxy variable is one of the proxy variables used to create the SR latent variable. The SR proxy variable has been used to measure self-regulation by Beaver, Ratchford and Ferguson (2009) and Vaughn et al. (2009). They report a value for Cronbach's alpha of .80 which indicates a high degree of correlation in the responses to the 20 questions. Miller, Barnes and Beaver (2011) examine the link between low values from the self-regulation proxy during adolescence and health problems in early adulthood. They use the Add Health self-regulation proxy and examined the relationship between varying levels of self-regulation and the likelihood of being diagnosed with a list of 10 physical and mental health conditions. They find that subjects with lower levels of self-regulation had significantly higher odds of being diagnosed with 9 of 10 negative health domains. Because the self-regulation proxy variable is the same value for each individual regardless of domains, it forces self-regulation to be a personality trait which is not generally supported by prior studies (Foxall et al., 2011, Jimura et al., 2011, Odum 2011). However, this measure is a good proxy variable for the self-regulation latent variable and provides an option to examine the relative explanatory power of the self-regulation latent variable.

Production efficiency in self-regulation has an inherited component. Timberlake et al. (2006) studied the association between smoking behavior and genes with wave 3 data from the Add Health. One gene variant was inversely associated with smoking accounting for approximately 1% of the variance in smoking prevalence. Never smokers and current nonsmokers had an excess of another gene variant compared with regular smokers, suggesting a protective effect. In another study by Hopfer et al. (2005) the association between drinking behavior and a set of gene variants was investigated. They used Add Health data and concluded that the gene variants accounted for from 3.1% to 2.0% of the variation in drinking behavior. Vaughn et al. (2009) also used the Add Health data and found a relationship between genetic variables and substance use.

Another psychological variable relates to childhood. Wright et al. (2012), Belsky and Beaver (2011) and Beaver, Ratchford and Ferguson (2009) show that some children exposed to difficult family environments later manifest problem behavior associated with low self-regulation. A maternal attachment scale was created from questions in wave1 of the Add Health. Because these data are were collected in wave 1 they can be considered retrospective in wave 3 and wave 4.

According to Baumeister et al. (2006) the exercise of self-regulation can enhance the individual's ability to self-regulate. Thus variables associated with greater exercise of self-regulation are likely to measure both greater levels of self-regulation and increased production efficiency. The psychological trait known as conscientiousness is one such variable. Conscientiousness is one of the big five personality traits and is defined as a tendency to show self-discipline, act dutifully and aim for achievement. The trait correlates with a preference for planned rather than spontaneous behavior. Jensen-Campbell et al. (2007) examined whether this personality trait is associated with the ability to exhibit self-regulation and moderate the anger–aggression link. Their results replicated previous findings that conscientiousness is negatively associated with anger and relative left prefrontal asymmetry. Conscientiousness was also found to moderate the link between anger and aggression. Individuals with higher scores on Conscientiousness are also more likely to have higher levels of education and health. Also individuals with ADHD are likely to score lower on conscientiousness.

The diversion of energy from self-regulation to competing psychological processes will also reduce self-regulation. That is, mental health issues can divert energy away from the production of self-regulation. Saffer and Dave (2005) show that mental health issues increase tobacco and alcohol use. Wave 4 of the Add Health includes three mental health variables which could affect substance use. They are measures of stress, depression and anxiety.

Religious adherence may also reflect the practice of self-regulation. All religions encourage the individual to avoid various temptations. McCullough and Willoughby (2009) review a number of studies which show that religious individuals are less likely to pursue unhealthy behaviors and have lower mortality than their less religious counterparts.

The production of self-regulation may also be affected by Attention Deficit Hyperactivity Disorder (ADHD). Past studies have found that that ADHD is correlated with a lack of ability to delay gratification. Wright et al. (2012) argue that ADHD, impulse regulation, the ability to delay gratification, and response inhibition are all highly heritable. Thus it is difficult to isolate a causal effect of ADHD that is independent of genetic measures. Parental socialization of children provides another route for the intergenerational transmission of these traits. Wave 3 of the Add Health contains 18 questions which measure retrospective hyperactivity. Following Vaughn et al. (2009) these questions were summed to form a non-clinical ADHD index.

Although behaviors associated with risk tolerance resemble behaviors associated with a limited ability to self-regulate, risk tolerance is probably not a good proxy for self-regulation.

There is evidence that smokers are aware of the risks of smoking yet continue to smoke (Viscusi, 1999 and Steptoe et al., 2002). In addition, studies have found that compared to nonsmokers, smokers are more likely to partake in a variety of other risky behaviors. For example, smokers tend to be more involved in traffic accidents (DiFranza et al. 1986), are less likely to wear seatbelts (Dillow, et al., 1981 and Eiser et al., 1997), and are more likely to engage in risky sexual behavior (Valois et al., 1999). These facts suggest that smokers are more risk tolerant than non-smokers. However, Ert, Yechiam and Arshavsky (2013), report on experimental studies which compared smokers to nonsmokers but do not find significant differences between the two groups in risk preference. If this is true, why do smokers make risky choices in various domains? Ert, Yechiam and Arshavsky (2013), using experimental data derived with the Iowa Gambling Task, conclude that the reason is a limited ability to self-regulate rather than risk preference. This makes risk preference and self-regulation distinctly different phenomena.

4.3 Other variables

Education is an important variable because it may be a proxy for health knowledge or efficiency in health production. Education is also likely to be correlated with self-regulation since individuals who have greater self-regulation have been shown to achieve higher levels of educational success (Mischel, Shoda and Peake, 1988). Education has also been shown to de-emphasize the role of emotion-based processes in decision making (Evans, Kemish and Turnbull, 2004). Education is not included as a proxy for self-regulation so that the effect of self-regulation on the education-health gradient can be measured and compared with results from prior studies of non-cognitive skills and the health-education gradient. Education is measured as years of schooling completed in wave 4.

The Add Health also includes a large number of questions which were used to construct demographic and economic variables. The Add Health demographic data includes gender, age, race, Hispanic, income, marital status and number of children. These variables were defined as either dichotomous or continuous. Table 1 includes the means and standard errors of all the variables including all the proxy variables.

5. Results

The first step in the empirical work is the estimation of equation (2) and the calculation of the domain specific self-regulation variables for all 11 domains as described in equation (3) and equation (4). The proxy variable coefficients from these regressions are presented in Appendix table 1 and are the empirical versions of equation (2). These empirical models also include education and demographic variables and fixed effects. The proxy variable coefficients are presented in order to highlight the differences in these coefficients across the domains.

Table 2 presents the bivariate correlations between the 11 domains. The reason for presenting this table is to contrast the correlations in domains with the correlations in SR in each domain. If the correlations across domains are high then it would be expected that the correlations in SR across domains would also be high. However, the results indicate that the correlations across domains are relatively small. These correlations are generally around .15. The strongest correlations for domains are between crime and substance use which are about .20. Table 2 also shows that the BMI and obesity variables have the smallest positive correlations, insignificant correlations and negative correlations. These correlations suggest that the choice of domains that individuals choose to engage in are not strongly related.

Table 3 presents the bivariate correlations between the estimated self-regulation latent variables calculated for each domain. Because each latent variable is a linear function of the proxy variables, some degree of correlation is expected. High correlations will be defined as over .75 and suggest that self-regulation in the pair of domains is similar. Table 3 shows, in contrast to the correlations between the domains, that the correlations in self-regulation across the11 domains are high. These correlations are generally around .85 which suggests that SR^{PH} across domains, including non-health domains, are similar. The exception to this is for BMI and obesity which have a low correlation with all the other domains. The correlations for the self-regulation BMI and obesity variables with the other self-regulation variables are generally around .30. The results from both table 2 and 3 show that excess eating is a unique domain with little relation to the other domains.

Table 4 presents the mean values for each of the estimated domain specific SR latent variables. These SR variables are all computed with the same choice of proxy variable 1 which gives them all the same metric. The table shows that BMI and obesity have the lowest levels of SR, cigarette consumption (by smokers only) is not much higher. The highest levels of SR are for marijuana use (for all individuals) and (for all individuals).

The next empirical task is to measure effect of the inclusion of SR^{PH} on health choices. Three specifications of equation (5) were estimated for each of the 11 domains and are presented in tables 5 through 9. The first specification does not include any proxy variables or any self-regulation measures. This specification establishes baseline values for the effects of demographics and the explanatory power of all the independent variables excluding SR. The second specification includes all the baseline specification variables and in addition includes the self-regulation proxy variable from the Add Health data set. This specification provides an option to compare the effects of the simple self-regulation proxy measure with the latent variable measure of self-regulation provided by the LW approach. The third specification includes all the baseline variables and includes the latent variable measure of self-regulation. Tables 5 through 9 provide an opportunity to compare the education coefficients in the baseline specification with the specification which includes the self-regulation latent variables. This comparison illustrates the effect of self-regulation on the education gradients.

Tables 5 through 9 show that self-regulation has a negative effect on all the domains and that the self-regulation latent variables have larger coefficients than the corresponding self-regulation proxy variables.[8] The self-regulation latent variables also increase the explanatory power of the empirical equation with respect to both the baseline specification and with respect to the simple proxy self-regulation specification. These tables also show that the inclusion of the self-regulation latent variable reduces the education gradient for smoking, drinking, drugs and crime but not for eating and gambling. In general, the gradients for black, Hispanic and female are not affected by the inclusion of the self-regulation latent variable except for eating where the effect of being female increases when self-regulation is included. For crime and gambling the inclusion of the self-regulation latent variable reduces the effect of being female.

6. Conclusions

Although it is not possible to definitively state that the self-regulation latent variables created in this study measure domain specific self-regulation, there is both theory and empirical evidence which support this interpretation. These self-regulation latent variables are novel, even though the proxy variables used to create these latent variables are familiar from prior studies. These prior studies provide both theoretical and empirical evidence that the proxy variables are related to self-regulation. The self-regulation latent variables are weighted averages of these proxy variables. Also, the empirical results in this study show that the self-regulation latent variables have significant and negative effects on domains which require self-regulation.

[8] Another variant of self-regulation is presented in Appendix table 1. This is the effect of the proxy self-regulation variable holding all the other proxies constant. These coefficients are similar to specification (2) of tables 5-9 but generally smaller in magnitude.

The empirical results suggest that there is a high correlation in self-regulation for smoking, drinking, drug use, crime and gambling, but that self-regulation for BMI and obesity are different from the other domains studied. That is, individuals with generally high self-regulation in smoking, drinking, drug use, crime and gambling may have low self-regulation in controlling their overeating. One possible explanation for this difference comes from evolutionary theory (Burnham, 2013). From an evolutionary perspective, if food shortages were the norm, then a preference for eating in excess, when possible, would favor survival. Thus, natural selection would result in an inclination to overeat when possible. However, a preference for all the other domains must be learned through experience. This could make self-regulation in eating different than self-regulation in the learned domains. The very low levels of SR in BMI and obesity highlight the need for increased public policies to reduce overeating.

The inclusion of the self-regulation latent variable reduces the effect of education except for marijuana use and eating. However, education is shown to retain a negative and significant effect on the domains which implies that education has a positive effect on health. This result supports the theory that education enhances efficiency in health production (Grossman, 2008).

Finally, the research presented in this paper raises questions about the effects of public policy on health related domains. Price, advertising and other policies may be moderated by interaction with self-regulation. Future research on how individual variations in SR affects price and advertising elasticities could be valuable in crafting new policies to promote health.

References

Baumeister, Roy F., Matthew Gailliot, C. Nathan DeWall, and Megan Oaten. "Self-Regulation and Personality: How Interventions Increase Regulatory Success, and How Depletion Moderates the Effects of Traits on Behavior." *Journal of Personality* 74, no. 6 (2006): 1773-1802.

Beaver, Kevin M., Marie Ratchford, and Christopher J. Ferguson. "Evidence of genetic and environmental effects on the development of low self-control." *Criminal Justice and Behavior* 36, no. 11 (2009): 1158-1172.

Becker, Gary S., and Casey B. Mulligan. "The endogenous determination of time preference." *The Quarterly Journal of Economics* 112, no. 3 (1997): 729-758.

Belsky, Jay, and Kevin M. Beaver. "Cumulative-genetic plasticity, parenting and adolescent self-regulation." *Journal of Child Psychology and Psychiatry* 52, no. 5 (2011): 619-626.

Burnham, Terence C. "Toward a neo-Darwinian synthesis of neoclassical and behavioral economics." *Journal of Economic Behavior & Organization* 90 (2013): S113-S127.

Casey, B. J., Adriana Galvan, and Todd A. Hare. "Changes in cerebral functional organization during cognitive development." *Current Opinion in Neurobiology* 15, no. 2 (2005): 239-244.

Charness, Gary, and Uri Gneezy. "Incentives to exercise." *Econometrica* 77, no. 3 (2009): 909-931.

Chiteji, Ngina. "Time-preference, Non-cognitive Skills and Well-being across the Life Course: Do Non-cognitive Skills Encourage Healthy Behavior?" *The American Economic Review* 100, no. 2 (2010): 200.

Conti, Gabriella, and Christopher Hansman. "Personality and the education–health gradient: A note on “Understanding differences in health behaviors by education”." *Journal of Health Economics* 32, no. 2 (2013): 480-485.

Conti, Gabriella, and James J. Heckman. "Understanding the Early Origins of the Education–Health Gradient A Framework That Can Also Be Applied to Analyze Gene–Environment Interactions." *Perspectives on Psychological Science* 5, no. 5 (2010): 585-605.

Conti, Gabriella, James Heckman, and Sergio Urzua. "The education-health gradient." *The American Economic Review* 100, no. 2 (2010): 234-238.

Cunha, Flavio, and James J. Heckman. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources* 43, no. 4 (2008): 738-782.

Cutler, David M., and Adriana Lleras-Muney. "Understanding differences in health behaviors by education." *Journal of Health Economics* 29, no. 1 (2010): 1-28.

Deckersbach, T., S. K. Das, L. E. Urban, T. Salinardi, P. Batra, A. M. Rodman, A. R. Arulpragasam, D. D. Dougherty and S. B. Roberts, "Pilot randomized trial demonstrating reversal of obesity-related abnormalities in reward system responsivity to food cues with a behavioral intervention", *Nutrition & Diabetes* (2014), 4, e129; doi:10.1038/nutd.2014.26.

DiFranza JR, Winters TH, Goldberg RJ, Cirillio L, Bilioris T (1986) "The relationship of smoking to motor vehicle accidents and traffic violations." New York State J Med 86: 464–7.

Dillow I, Swann C, Cliff KS (1981), "A study of the effect of a health education programme in promoting seat-belt wearing", *Health Education J* 40: 14–8. doi:10.1177/001789698104000107.

Eiser JR, Sutton SR, Wober M (1997) Smoking, seatbelts, and beliefs about health. *Addictive Behavior* 4: 331–8.

Ert, Eyal, Eldad Yechiam, and Olga Arshavsky. "Smokers' Decision Making: More than Mere Risk Taking." *PloS one* 8, no. 7 (2013): e68064.

Evans, Cathryn EY, Karen Kemish, and Oliver H. Turnbull. "Paradoxical effects of education on the Iowa Gambling Task." *Brain and Cognition* 54, no. 3 (2004): 240-244.

Frederick, Shane, George Loewenstein, and Ted O'donoghue. "Time discounting and time preference: A critical review." *Journal of Economic Literature* 40, no. 2 (2002): 351-401.

Foxall, Gordon R., John R. Doyle, Mirella Yani-de-Soriano, and Victoria K. Wells. "Contexts and individual differences as influences on consumers' delay discounting." *Psychological Record* 61, no. 4 (2011).

Fudenberg, Drew, and David K. Levine. "A dual-self model of impulse control." *The American Economic Review* (2006): 1449-1476.

Gailliot, Matthew T., and Roy F. Baumeister. "The physiology of willpower: Linking blood glucose to self-control." *Personality and Social Psychology Review* 11, no. 4 (2007): 303-327.

Grossman, Michael. "The relationship between health and schooling." *Eastern Economic Journal* 34, no. 3 (2008): 281-292.

Gubler, Timothy, and Lamar Pierce. "Healthy, Wealthy, and Wise: Retirement Planning Predicts Employee Health Improvements." (2014) http://apps.olin.wustl.edu/faculty/pierce/401k-paper_rev4.pdf

Heckman, James J. "The case for investing in disadvantaged young children." *Big ideas for children: Investing in our nation's future.* (2008): 49-58.

Hopfer, Christian J., David Timberlake, Brett Haberstick, Jeffrey M. Lessem, Marissa A. Ehringer, Andrew Smolen, and John K. Hewitt. "Genetic influences on quantity of alcohol consumed by adolescents and young adults." *Drug and Alcohol Dependence* 78, no. 2 (2005): 187-193.

Jensen-Campbell, Lauri A., Jennifer M. Knack, Amy M. Waldrip, and Shaun D. Campbell. "Do Big Five personality traits associated with self-control influence the regulation of anger and aggression?" *Journal of Research in Personality* 41, no. 2 (2007): 403-424.

Jimura, Koji, Joel Myerson, Joseph Hilgard, Julia Keighley, Todd S. Braver, and Leonard Green. "Domain independence and stability in young and older adults' discounting of delayed rewards." *Behavioural Processes* 87, no. 3 (2011): 253-259.

Kaestner Robert and Kevin Callison. "Adolescent Cognitive and Noncognitive Correlates of Adult Health*."Journal of Human Capital* , Vol. 5, No. 1 (Spring 2011), pp. 29-69

Lubotsky, Darren, and Martin Wittenberg. "Interpretation of regressions with multiple proxies." *The Review of Economics and Statistics* 88, no. 3 (2006): 549-562.

MacKillop, James, and Christopher W. Kahler. "Delayed reward discounting predicts treatment response for heavy drinkers receiving smoking cessation treatment." *Drug and Alcohol Dependence* 104, no. 3 (2009): 197-203.

McCullough, Michael E., and Brian LB Willoughby. "Religion, self-regulation, and self-control: Associations, explanations, and implications." *Psychological Bulletin* 135, no. 1 (2009): 69.

Miller, Holly Ventura, J. C. Barnes, and Kevin M. Beaver. "Self-control and health outcomes in a nationally representative sample." *American Journal of Health Behavior* 35, no. 1 (2011): 15-27.

Mischel, Walter, Yuichi Shoda, and Philip K. Peake. "The nature of adolescent competencies predicted by preschool delay of gratification." *Journal of Personality and Social Psychology* 54, no. 4 (1988): 687.

Moffitt, Terrie E., Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, HonaLee Harrington, Renate Houts et al. "A gradient of childhood self-control predicts health, wealth, and public safety." *Proceedings of the National Academy of Sciences* 108, no. 7 (2011): 2693-2698.

Odum, Amy L. "Delay discounting: Trait variable?" *Behavioural Processes* 87, no. 1 (2011): 1-9.

Saffer, Henry, and Dhaval Dave. "Mental illness and the demand for alcohol, cocaine, and cigarettes." *Economic Inquiry* 43, no. 2 (2005): 229-246.

Savelyev, Peter A. "Psychological Skills, Education, and Longevity of High-Ability Individuals." Unpublished manuscript, Vanderbilt University, Department of Economics (2014).

Stanovich, Keith E., and Richard F. West. "On the relative independence of thinking biases and cognitive ability." *Journal of Personality and Social Psychology* 94, no. 4 (2008): 672.

Steptoe A, Wardle J, Weiwei C, Baban A, Glass K, et al. (2002) An international comparison of tobacco smoking, beliefs, and risk awareness in university students from 23 countries. *Addiction* 97: 1561–71.

Timberlake, David S., Brett C. Haberstick, Jeffrey M. Lessem, Andrew Smolen, Marissa Ehringer, John K. Hewitt, and Christian Hopfer. "An association between the DAT1 polymorphism and smoking behavior in young adults from the National Longitudinal Study of Adolescent Health." *Health Psychology* 25, no. 2 (2006): 190.

Valois RF, Oeltmann JE, Waller J, Hussey JR (1999) Relationship between number of sexual intercourse partners and selected health risk behaviors among public high school adolescents. *J Adol Health* 25: 328–35.

Vaughn, Michael G., Kevin M. Beaver, Matt DeLisi, Brian E. Perron, and Lisa Schelbe. "Gene-environment interplay and the importance of self-control in predicting polydrug use and substance-related problems." *Addictive Behaviors* 34, no. 1 (2009): 112-116.

Viscusi WK (1999). Do smokers underestimate risks? J Polit Econ 98: 1253–69.

Wright, John Paul, Rebecca Schnupp, Kevin M. Beaver, Matt Delisi, and Michael Vaughn. "Genes, Maternal Negativity, and Self-Control Evidence of a Gene× Environment Interaction." *Youth Violence and Juvenile Justice* 10, no. 3 (2012): 245-260.

Table 1: Summary Statistics

	Mean	SD	Min	Max
Outcomes				
Smoker	0.37	0.48	0	1
Cigarette Consumption of Smokers per month	318.50	305.35	0	3000
Alcohol Consumption of Drinkers per month	26.83	48.63	0	540
Regular Binge Drinker	0.28	0.45	0	1
Marijuana Use in Past 30 Days	2.32	9.57	0	200
Num of Drug Abuse Symptoms	0.14	0.58	0	4
Ever Been Arrested	0.21	0.41	0	1
Index of Criminal Behavior and Violence	0.39	0.96	0	10
Gambler	0.04	0.19	0	1
Obese	0.30	0.46	0	1
BMI	27.71	6.94	13	72
Self-Regulation Proxies				
Self-Regulation Index	65.71	7.92	38	96
Genetic Index	1.59	0.91	0	4
Maternal Attachment	9.20	1.18	2	10
Conscientiousness Scale	14.55	2.70	4	20
Perceived Stress Scale	4.82	2.98	0	16
CESD Depression Scale	2.46	2.58	0	15
Anxious Personality Scale	12.33	2.96	4	20
Religious	0.50	0.50	0	1
ADHD index	13.65	9.05	0	54
Demographic Characteristics				
White	0.67	0.47	0	1
Black	0.16	0.37	0	1
Hispanic	0.12	0.32	0	1
Asian	0.03	0.18	0	1
Native American	0.01	0.09	0	1
Other Race	0.01	0.09	0	1
Female	0.49	0.50	0	1
Age	25.09	3.77	18	34
Age Squared	643.84	189.88	324	1156
Years of Education (by W4)	14.15	2.21	8	20
Earnings in $10000	2.39	3.45	0	100
Married	0.29	0.45	0	1
Number of Children in the HH	0.57	0.96	0	7

Observations: 29122

Table 2: Correlation Across Outcomes

	Smoker	Cigarettes	Alcohol Consum.	Binge Drinker	Marijuana	Drug Abuse	Arrested	Crime Index	Gambler	Obese	BMI
SMOKER	1										
CIGARETTES		1									
ALCOHOL	.197	.124	1								
BINGE DRINKER	.222	.049	.504	1							
MARIJUANA	.154	.064	.142	.145	1						
DRUG ABUSE	.173	.082	.124	.116	.123	1					
ARRESTED	.238	.070	.182	.182	.103	.200	1				
CRIME INDEX	.151	.046	.228	.195	.271	.190	.228	1			
GAMBLER	.075	.045	.103	.085	.046	.063	.094	.096	1		
OBESE	.001	.006	-.027	-.044	-.042	-.030	.012	-.026	.032	1	
BMI	-.018	-.013	-.028	-.048	-.058	-.039	.012	-.027	.038	.808	1

Description: All Pair wise correlations between outcomes of interest. All correlation coefficients are significant at the 5% level except for the correlations between obese and smoker, obese and cigarettes, and BMI and cigarettes.

Table 3: Correlation Across Self-Regulation Skill Latent Variable Measures

	Smoker	Cigarettes	Alcohol Consum.	Binge Drinker	Marijuana	Drug Abuse	Arrested	Crime Index	Gambler	Obese	BMI
Smoker	1										
Cigarettes	.897	1									
Alcohol Con	.900	.758	1								
Binge Drinker	.903	.692	0.951	1							
Marijuana	.924	.700	0.924	.969	1						
Drug Abuse	.906	.846	0.868	.834	0.781	1					
Arrested	.944	.828	0.956	.915	0.907	.897	1				
Crime Index	.859	.773	0.913	.884	0.824	.939	.897	1			
Gambler	.788	.631	0.911	.846	0.836	.797	.814	.820	1		
Obese	**.239**	**.381**	**0.254**	**.154**	**0.180**	**.345**	**.134**	**.382**	**.492**	1	
BMI	**.250**	**.368**	**0.300**	**.179**	**0.203**	**.355**	**.160**	**.412**	**.537**	.986	1

Description: All Pair wise correlations between the constructed self-regulation proxies. All of these correlations are statistically significant at the 5% level.

Table 4

SRS^{PH} Latent Variable for:	Mean
Cigarette Consumption	4.737
Smoker	24.503
Alcohol Consumption	40.078
Binge	55.013
Marijuana	64.906
Drug Abuse	27.273
Obesity	0.0816
BMI	1.4939
Arrested	27.945
Crime	41.001
Gambler	36.940

Table 5
Effect of Self-Regulation on Smoking

	SMOKER	SMOKER	SMOKER	CIGARETTES	CIGARETTES	CIGARETTES
BLACK	-0.152***	-0.147***	-0.138***	-153.920***	-152.620***	-141.656***
	(0.014)	(0.014)	(0.009)	(18.378)	(19.578)	(11.895)
HISPANIC	-0.137***	-0.126***	-0.118***	-134.434***	-127.222***	-122.420***
	(0.019)	(0.022)	(0.010)	(19.386)	(19.728)	(11.991)
ASIAN	-0.052	-0.060*	-0.046***	-104.886***	-98.669***	-85.062***
	(0.038)	(0.034)	(0.014)	(14.676)	(13.835)	(15.200)
NATIVE AMERICAN	0.041	0.028	-0.002	-19.918	4.136	14.148
	(0.046)	(0.046)	(0.039)	(41.166)	(45.642)	(43.491)
OTHER	-0.062	-0.064	-0.054	-47.379	-40.841	-30.251
	(0.044)	(0.044)	(0.036)	(30.214)	(35.881)	(38.766)
FEMALE	-0.049***	-0.025*	-0.021***	-58.379***	-51.514***	-60.723***
	(0.010)	(0.013)	(0.007)	(5.556)	(6.347)	(7.008)
AGE	0.029***	0.024*	0.016	32.322**	31.160**	22.428
	(0.010)	(0.012)	(0.013)	(12.513)	(11.692)	(13.903)
AGE SQUARED	-0.001***	-0.000*	-0.000	-0.541**	-0.498**	-0.342
	(0.000)	(0.000)	(0.000)	(0.242)	(0.236)	(0.270)
EDUCATION (BY W4)	-0.054***	-0.052***	-0.046***	-26.126***	-26.902***	-22.608***
	(0.003)	(0.004)	(0.002)	(2.065)	(2.208)	(2.054)
EARNINGS IN $10000	-0.002***	-0.002*	-0.001	-1.027	-1.018	-1.131
	(0.001)	(0.001)	(0.001)	(0.723)	(0.786)	(0.818)
MARRIED	-0.128***	-0.114***	-0.102***	-21.982**	-18.262*	-6.629
	(0.007)	(0.008)	(0.008)	(9.869)	(9.816)	(8.748)
CHILDREN IN THE HH	0.011**	0.017**	0.018***	10.155**	11.463***	9.544**
	(0.005)	(0.007)	(0.004)	(3.939)	(4.080)	(4.001)
SELF-REG. PROXY		-0.008***			-2.052***	
		(0.001)			(0.482)	
SELF-REG. LATENT			-0.018***			-13.090***
			(0.001)			(1.203)
OBSERVATIONS	26,162	22,097	20,216	9,021	7,515	6,844
R-SQUARED	0.119	0.137	0.153	0.123	0.128	0.142

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Standard Errors Clustered at the State Level. State and Time Fixed Effects are included in all regressions.

Table 6
Effect of Self-Regulation on Drinking

	Alcohol Consumption	Alcohol Consumption	Alcohol Consumption	Binge Drinker	Binge Drinker	Binge Drinker
Black	-8.716***	-8.359***	-7.628***	-0.144***	-0.142***	-0.136***
	(1.271)	(1.267)	(1.172)	(0.009)	(0.009)	(0.008)
Hispanic	-7.582***	-6.556***	-5.754***	-0.068***	-0.056***	-0.055***
	(1.248)	(1.523)	(1.222)	(0.018)	(0.019)	(0.010)
Asian	-11.493***	-11.950***	-12.113***	-0.108***	-0.116***	-0.112***
	(1.335)	(1.685)	(1.388)	(0.021)	(0.024)	(0.013)
Native American	1.585	0.092	0.960	0.040	-0.001	0.009
	(3.183)	(3.904)	(4.809)	(0.032)	(0.036)	(0.038)
Other	-9.834***	-8.640**	-8.944**	-0.115***	-0.099**	-0.092***
	(3.051)	(3.806)	(3.645)	(0.037)	(0.041)	(0.032)
Female	-19.510***	-17.552***	-17.280***	-0.136***	-0.119***	-0.116***
	(1.246)	(1.332)	(0.767)	(0.008)	(0.009)	(0.006)
Age	-1.842	-1.156	-1.825	0.000	0.011	0.007
	(1.602)	(1.515)	(1.706)	(0.011)	(0.012)	(0.013)
Age Squared	0.033	0.025	0.036	-0.000	-0.000	-0.000
	(0.030)	(0.028)	(0.033)	(0.000)	(0.000)	(0.000)
Years of Education (W4)	-1.732***	-1.678***	-1.239***	-0.007***	-0.007***	-0.004***
	(0.220)	(0.243)	(0.188)	(0.001)	(0.002)	(0.001)
Earnings in $10000	-0.048	-0.021	0.004	0.002**	0.002**	0.002**
	(0.066)	(0.064)	(0.077)	(0.001)	(0.001)	(0.001)
Married	-9.571***	-8.597***	-8.188***	-0.114***	-0.108***	-0.104***
	(0.906)	(1.085)	(0.845)	(0.008)	(0.008)	(0.007)
Children in HH	-2.102***	-1.862***	-1.871***	-0.017***	-0.012***	-0.010***
	(0.443)	(0.442)	(0.477)	(0.003)	(0.004)	(0.003)
Self-Regulation Proxy		-0.761***			-0.007***	
		(0.051)			(0.000)	
Self-Regulation Latent			-1.002***			-0.008***
			(0.065)			(0.000)
Observations	19,139	16,415	15,071	26,336	22,175	20,282
R-squared	0.080	0.093	0.100	0.088	0.106	0.111

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Standard Errors Clustered at the State Level. State and Time Fixed Effects are included in all regressions.

Table 7
Effect of Self-Regulation on Drug Use

	MARIJUANA	MARIJUANA	MARIJUANA	DRUG ABUSE	DRUG ABUSE	DRUG ABUSE
BLACK	-0.062	0.110	0.214	-0.179***	-0.172***	-0.168***
	(0.219)	(0.234)	(0.206)	(0.015)	(0.016)	(0.010)
HISPANIC	-0.709***	-0.659**	-0.743***	-0.121***	-0.111***	-0.104***
	(0.249)	(0.302)	(0.202)	(0.011)	(0.012)	(0.014)
ASIAN	-1.251***	-1.400***	-1.312***	-0.135***	-0.134***	-0.125***
	(0.213)	(0.213)	(0.281)	(0.018)	(0.017)	(0.017)
NATIVE AMERICAN	-0.614	-1.039**	-0.980*	-0.090	-0.084	-0.099*
	(0.380)	(0.441)	(0.560)	(0.081)	(0.085)	(0.060)
OTHER	-0.221	-0.037	0.345	-0.099	-0.083	-0.085*
	(0.593)	(0.713)	(0.998)	(0.070)	(0.075)	(0.045)
FEMALE	-1.423***	-1.166***	-0.973***	-0.039***	-0.016	-0.027***
	(0.157)	(0.162)	(0.143)	(0.009)	(0.011)	(0.009)
AGE	-0.803***	-0.733***	-0.694**	0.035***	0.030***	0.033*
	(0.273)	(0.270)	(0.306)	(0.009)	(0.008)	(0.017)
AGE SQUARED	0.013***	0.013***	0.012**	-0.001***	-0.000***	-0.001
	(0.005)	(0.005)	(0.005)	(0.000)	(0.000)	(0.000)
EDUCATION (BY W4)	-0.293***	-0.326***	-0.284***	-0.022***	-0.021***	-0.016***
	(0.036)	(0.043)	(0.035)	(0.004)	(0.004)	(0.002)
EARNINGS IN $10000	-0.018**	-0.017*	-0.013	-0.004***	-0.003***	-0.003***
	(0.008)	(0.010)	(0.010)	(0.001)	(0.001)	(0.001)
MARRIED	-0.869***	-0.693***	-0.493***	-0.094***	-0.073***	-0.065***
	(0.102)	(0.116)	(0.112)	(0.009)	(0.010)	(0.010)
CHILDREN IN HH	-0.106**	-0.098	-0.089	-0.003	0.001	-0.000
	(0.049)	(0.062)	(0.059)	(0.004)	(0.005)	(0.005)
SELF-REG. PROXY		-0.126***			-0.008***	
		(0.014)			(0.001)	
SELF-REG. LATENT			-0.138***			-0.015***
			(0.010)			(0.001)
OBSERVATIONS	24,401	20,686	18,938	22,876	19,286	17,658
R-SQUARED	0.049	0.060	0.062	0.037	0.051	0.061

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Standard Errors Clustered at the State Level. State and Time Fixed Effects are included in all regressions.

Table 8
Effect of Self-Regulation on Criminal Behavior and Gambling

	ARRESTED	ARRESTED	ARRESTED	CRIME INDEX	CRIME INDEX	CRIME INDEX	GAMBLER	GAMBLER	GAMBLER
BLACK	0.013	0.012	0.022***	0.072***	0.093***	0.091***	-0.006	-0.005	-0.004
	(0.009)	(0.008)	(0.008)	(0.024)	(0.026)	(0.020)	(0.005)	(0.005)	(0.003)
HISPANIC	-0.037***	-0.026**	-0.020**	-0.016	0.002	0.008	-0.008	-0.005	-0.002
	(0.010)	(0.011)	(0.008)	(0.016)	(0.016)	(0.020)	(0.007)	(0.007)	(0.004)
ASIAN	-0.064***	-0.063***	-0.056***	0.002	-0.012	-0.019	0.039***	0.039***	0.041***
	(0.011)	(0.008)	(0.011)	(0.017)	(0.022)	(0.028)	(0.008)	(0.009)	(0.008)
NATIVE AMERICAN	0.077**	0.061	0.049	0.123	0.132	0.099	-0.018*	-0.019	-0.015
	(0.038)	(0.041)	(0.033)	(0.073)	(0.079)	(0.083)	(0.011)	(0.012)	(0.013)
OTHER	-0.035	-0.023	-0.008	-0.097*	-0.082	-0.054	0.018	0.015	0.026
	(0.036)	(0.040)	(0.029)	(0.048)	(0.067)	(0.059)	(0.027)	(0.027)	(0.020)
FEMALE	-0.186***	-0.162***	-0.161***	-0.355***	-0.308***	-0.328***	-0.049***	-0.044***	-0.040***
	(0.007)	(0.007)	(0.005)	(0.017)	(0.015)	(0.013)	(0.006)	(0.005)	(0.003)
AGE	0.013	0.013	0.014	-0.194***	-0.210***	-0.209***	0.008**	0.008**	0.007
	(0.009)	(0.009)	(0.010)	(0.019)	(0.023)	(0.030)	(0.003)	(0.003)	(0.005)
AGE SQUARED	-0.000	-0.000	-0.000	0.003***	0.003***	0.003***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
EDUCATION (BY W4)	-0.030***	-0.029***	-0.025***	-0.031***	-0.030***	-0.020***	-0.004***	-0.004***	-0.003***
	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
EARNINGS IN $10000	-0.003***	-0.002**	-0.002**	-0.008***	-0.006***	-0.006***	0.001***	0.001***	0.001***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
MARRIED	-0.100***	-0.088***	-0.078***	-0.156***	-0.116***	-0.092***	-0.011***	-0.009***	-0.010***
	(0.007)	(0.008)	(0.006)	(0.013)	(0.014)	(0.013)	(0.003)	(0.003)	(0.003)
CHILDREN IN HH	0.009***	0.010***	0.009**	-0.019***	-0.014*	-0.019***	0.000	0.000	0.001
	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)	(0.002)	(0.002)	(0.001)
SELF-REG. PROXY		-0.006***			-0.019***			-0.001***	
		(0.001)			(0.001)			(0.000)	
SELF-REG. LATENT			-0.011***			-0.027***			-0.002***
			(0.001)			(0.001)			(0.000)
OBSERVATIONS	26,177	22,101	20,219	24,642	20,920	19,145	23,770	22,175	20,282
R-SQUARED	0.150	0.157	0.166	0.086	0.115	0.129	0.037	0.041	0.041

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Standard Errors Clustered at the State Level. State and Time Fixed Effects are included in all regressions.

Table 9
Effect of Self-Regulation on Obesity

	OBESE	OBESE	OBESE	BMI	BMI	BMI
BLACK	0.084***	0.080***	0.083***	1.672***	1.585***	1.593***
	(0.011)	(0.013)	(0.010)	(0.200)	(0.207)	(0.152)
HISPANIC	0.063***	0.071***	0.068***	1.179***	1.258***	1.259***
	(0.013)	(0.018)	(0.010)	(0.119)	(0.177)	(0.151)
ASIAN	-0.043	-0.043	-0.040***	-0.633	-0.577	-0.542***
	(0.031)	(0.033)	(0.013)	(0.561)	(0.595)	(0.201)
NATIVE AMERICAN	0.135	0.137	0.135***	3.216*	3.286*	3.016***
	(0.095)	(0.090)	(0.040)	(1.840)	(1.675)	(0.730)
OTHER	-0.030	-0.014	-0.016	-0.258	-0.064	-0.094
	(0.052)	(0.065)	(0.035)	(0.625)	(0.789)	(0.467)
FEMALE	0.019*	0.024**	0.036***	0.106	0.227	0.432***
	(0.010)	(0.011)	(0.007)	(0.165)	(0.152)	(0.102)
AGE	0.039***	0.043***	0.047***	0.872***	0.969***	1.078***
	(0.009)	(0.009)	(0.013)	(0.146)	(0.138)	(0.195)
AGE SQUARED	-0.001***	-0.001***	-0.001***	-0.013***	-0.015***	-0.017***
	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)	(0.004)
EDUCATION (BY W4)	-0.021***	-0.021***	-0.019***	-0.322***	-0.331***	-0.306***
	(0.003)	(0.002)	(0.002)	(0.048)	(0.048)	(0.023)
EARNINGS IN $10000	-0.002**	-0.003***	-0.002**	-0.034***	-0.045***	-0.046***
	(0.001)	(0.001)	(0.001)	(0.010)	(0.010)	(0.012)
MARRIED	0.021***	0.026***	0.029***	0.374***	0.410***	0.415***
	(0.007)	(0.007)	(0.008)	(0.109)	(0.117)	(0.122)
CHILDREN IN HH	0.020***	0.017***	0.015***	0.248***	0.211***	0.185***
	(0.004)	(0.004)	(0.004)	(0.063)	(0.075)	(0.064)
SELF-REG. PROXY		-0.001**			-0.026**	
		(0.000)			(0.010)	
SELF-REG. LATENT			-1.526***			-1.039***
			(0.111)			(0.065)
OBSERVATIONS	25,624	21,579	19,794	25,624	21,579	19,794
R-SQUARED	0.057	0.061	0.070	0.075	0.081	0.092

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Standard Errors Clustered at the State Level. State and Time Fixed Effects are included in all regressions.

Appendix 1: Detailed empirical definition of the self–regulation proxy variables

1) Self-regulation proxy; Data from wave 3; Definition: 20 questions relating to self-regulation, 17 of the questions employ a five point response scale (1-5) and three questions employ a four point (1-4) response scale. The four point response scale was rescaled to a five point range by recoding two as 2.25 and three as 3.75 and four as five. All of the responses were coded such that higher values correspond to high self-regulation. The self-regulation proxy ranges from 20 to 100 with higher values representing more self-regulation.

2) Genetic proxy; Data wave 4; Definition: the presence of the 0R allele of DAT1, the 7R allele of DRD4, the short allele of 5HTTLPR, and the 2R/3R alleles of MAOA. (The A1 allele of DRD2 is not included in the Wave 4 DNA data.) This index ranges from 0 to 4 with higher values indicating a lesser likelihood that genetic factors will influence the individual's behavior i.e. more self-regulation.

3) Maternal Attachment Proxy; Data from wave 1; Definition: Maternal Attachment and is based on two questions on the respondents feelings about their mother. The data has response codes from 1 to 5 creating a scale equal to 2 to 10 with higher values indicating more attachment.

4) Conscientiousness; Data from wave 4; Definition: Four questions which relate to orderliness scale from 4 to 20 with 4 the least conscientious and 20 the most conscientious. Add Health constructed variable: C4VAR007.

5) Stress; Data from wave 4; Definition: Cohen Perceived Stress Scale outcomes between 0 and 16 with higher values indicating more perceived stress. Add Health constructed variable: C4VAR001.

6) Depression; Data from wave 4; Definition: Center for Epidemiological Studies depression (CESD). A five question version of this scale has outcomes between 0 and 15 with higher values indicating greater depression. Add Health constructed variable: C4VAR002.

7) Anxiety; Data from wave 4; Definition: Anxious Personality Scale which has outcomes from 0-20 higher values indicating greater anxiousness. Add Health constructed variable: C4VAR009.

8) Religiosity; Data from wave 3; Definition: How often the respondent has attended a religious service in the past 12 months. Defined as a dichotomous variable equal to one if the respondent indicates that they attended religious services two or three times a month or more.

9) Attention Deficit Hyperactivity Disorder (ADHD); Data from wave 3; Definition: Based on 18 questions which measure retrospective hyperactivity. The response code for the questions is 0 through 3 which creates a proxy ranging from 0 to 54. The value of 0 represents the lowest probability of ADHD and 54 represents the highest probability of ADHD. This is a probabilistic non-clinical measure of ADHD.

Appendix 2:

Differences in the effects of the proxy variables for each outcome provide some insight into differences in self-regulation by outcome. The proxy coefficients presented in Appendix table 1 are described as negative or positive if they are significant. The self-regulation proxy variable is negative for all outcomes except eating. The genetic index variable is negative for obesity but and positive for drug abuse and the crime index. The maternal attachment index follows a similar pattern to the genetic proxy by being positive for eating and negative for crime. Conscientiousness is negative for eating and crime outcome but positive for the probability of arrest. Stress is positive for smoking, binge drinking, and drug use. Depression is positive for smoking, drinking, and crime. Anxiety does not follow any obvious pattern. Religiosity is negative except for eating where it is positive and it is insignificant for one crime outcome and for gambling. ADHD is positive for all outcomes. One important pattern in these results is that the effects of the proxies on the eating outcomes, obesity and BMI, are generally different from the other outcomes. Because the proxies are used to create the self-regulation variables, these results suggest that the process of self-regulation in eating is different than it is for the other outcomes.

Appendix Table 1
Coefficients of the Proxy Variables Only

	SMOKER	CIGARETTES	ALCOHOL CONSUMPTION	BINGE DRINKER	MARIJUANA	DRUG ABUSE
SELF-REGULATION PROXY	-0.007***	-1.001*	-0.674***	-0.007***	-0.114***	-0.007***
	(0.001)	(0.527)	(0.063)	(0.001)	(0.017)	(0.001)
GENETIC INDEX	0.003	-4.262	-0.579*	0.001	0.004	0.008*
	(0.003)	(3.908)	(0.308)	(0.004)	(0.070)	(0.004)
MATERNAL ATTACHMENT	-0.002	-1.401	-0.219	-0.002	-0.028	-0.009
	(0.003)	(2.675)	(0.374)	(0.002)	(0.059)	(0.006)
CONSCIENTIOUSNESS SCALE	-0.001	-1.882	0.362**	0.001	0.001	-0.003
	(0.002)	(1.228)	(0.140)	(0.001)	(0.033)	(0.002)
PERCEIVED STRESS SCALE	0.009***	3.599**	0.131	-0.001	0.035	0.008***
	(0.001)	(1.333)	(0.180)	(0.001)	(0.025)	(0.003)
CESD DEPRESSION SCALE	0.003*	2.884***	0.566***	0.005***	0.064	0.003
	(0.002)	(1.026)	(0.209)	(0.001)	(0.042)	(0.002)
ANXIOUS PERSONALITY SCALE	-0.001	1.220	-0.206	-0.000	-0.096***	0.006***
	(0.001)	(1.308)	(0.180)	(0.001)	(0.027)	(0.002)
RELIGIOUS	-0.091***	-38.221***	-3.094***	-0.046***	-1.080***	-0.025**
	(0.009)	(8.325)	(0.836)	(0.008)	(0.193)	(0.011)
ADHD INDEX	0.003***	2.387***	0.297***	0.001*	0.016**	0.003***
	(0.001)	(0.368)	(0.063)	(0.000)	(0.007)	(0.001)
OBSERVATIONS	20,216	6,844	15,071	20,282	18,898	17,658
R-SQUARED	0.153	0.142	0.100	0.111	0.062	0.061

Robust standard errors in parentheses. Standard errors clustered at the state level. Education and demographic variables listed in table 4 are also included in all regressions. State and time fixed effects are included in all regressions. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 1 continued

	Arrested	Crime Index	Gambler	Obese	BMI
SELF-REGULATION PROXY	-0.005***	-0.016***	-0.001***	-0.000	-0.012
	(0.001)	(0.001)	(0.000)	(0.001)	(0.011)
GENETIC INDEX	0.003	0.010	0.001	-0.006**	-0.111**
	(0.002)	(0.006)	(0.001)	(0.003)	(0.045)
MATERNAL ATTACHMENT	-0.010***	-0.012**	0.001	0.009***	0.226***
	(0.003)	(0.005)	(0.001)	(0.003)	(0.044)
CONSCIENTIOUSNESS SCALE	0.004***	-0.006**	-0.000	-0.013***	-0.199***
	(0.001)	(0.003)	(0.001)	(0.002)	(0.024)
PERCEIVED STRESS SCALE	0.004***	0.005**	-0.000	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.002)	(0.029)
CESD DEPRESSION SCALE	0.005***	0.031***	-0.000	0.001	0.032
	(0.001)	(0.003)	(0.001)	(0.001)	(0.023)
ANXIOUS PERSONALITY SCALE	-0.002	0.001	-0.000	-0.002	-0.051**
	(0.001)	(0.003)	(0.001)	(0.001)	(0.020)
RELIGIOUS	-0.040***	-0.025	-0.001	0.024**	0.516***
	(0.007)	(0.019)	(0.003)	(0.010)	(0.147)
ADHD INDEX	0.002***	0.004***	0.001**	0.002***	0.039***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.008)
OBSERVATIONS	20,219	19,145	20,282	19,794	19,794
R-SQUARED	0.166	0.129	0.041	0.070	0.092

Robust standard errors in parentheses. Standard errors clustered at the state level. Education and demographic variables listed in table 4 are also included in all regressions. State and time fixed effects are included in all regressions. *** p<0.01, ** p<0.05, * p<0.1