H I E R

Harvard Institute of Economic Research

Discussion Paper Number 2060

What Explains Differences in Smoking, Drinking and
Other Health-Related Behaviors?

by

David Cutler and Edward Glaeser

February 2005

Harvard University
Cambridge, Massachusetts

What Explains Differences in Smoking, Drinking and Other Health-Related Behaviors?

by

David Cutler and Edward Glaeser
Harvard University and NBER

January 2005

Abstract

We explore economic model of health behaviors. While the standard economic model of health as an investment is generally supported empirically, the ability of this model to explain heterogeneity across individuals is extremely limited. Most prominently, the correlation of different health behaviors across people is virtually zero, suggest that standard factors such as variation in discount rates or the value of life are not the drivers of behavior. We focus instead on two other factors: genetics; and behavioral-specific situational factors. The first factor is empirically important, and we suspect the second is as well.

Unhealthy behaviors are a major part of the health profile in the United States and other developed countries. A famous article by J. Michael McGinnis, and William H. Foege (1993) estimated that half of all deaths in the United States are due to unhealthy behaviors, most importantly tobacco use, poor diet and exercise, and excessive alcohol consumption. Even this excludes many health-harming behaviors such as passing up recommended screening tests, discontinuing medication usage, and not following up on physician referrals. Why do people not do the right thing for their health?

The prevailing economic paradigm for understanding these forms of behavior is that health behaviors are investments, where foregone current pleasure leads to improvements in future well-being (Michael Grossman, 1972). As such, variations in health behaviors will be explained by differences in discount rates or the value of life.

If this theory is right, health behaviors will differ across people, but for a given person, behaviors will be highly correlated: those who value their health highly and care sufficiently about the future will have much better behaviors than those who do not. Table 1 shows the correlation of five different health measures, drawn from the 1990 National Health Interview Survey (NHIS): current smoking; heavy drinking (3 or more drinks per day); being obese (BMI≥30); using hypertension medication when recommended; and (for women) receiving a mammogram in the past 3 years. In each case, we sample the population aged 45 and older, a group where the health consequences of poor behavior should be salient. The correlations across behaviors are surprisingly low. Smoking and drinking exhibit the largest correlation, but this is only 16 percent.[1] Most correlations are below 10 percent.

The low correlation is not an artifact of this data set or time period. Appendix Table 1 shows analysis of data from the Behavioral Risk Factor Surveillance Survey in 2002. This survey includes measures of smoking, drinking, obesity, use of a seatbelt, receipt of a flu shot in the past year, and screening for colorectal, breast, cervical, and prostate cancer. Preventive service receipt is more correlated than is dietary behavior, with roughly 20 percent correlations among these activities. But even 20 percent is not very high, and preventive service use is virtually unrelated to dietary behavior. The individual model of behavior is thus quantitatively not very important.

Even if health behaviors differ generally, individuals might at some point in their lives experience shocks to their discount rates, incomes, or beliefs about the future that will result in large behavioral changes. Thus, year to year changes in health behaviors could be more highly correlated than behavior at any point in time. We examine this proposition using data from the Health and Retirement Study, a sample of nearly 10,000 people nearing retirement age. There are six waves of data available to this point.

Table 2 shows correlations of changes in behavior. In each case, the sample is people who were not maximizing their health beforehand: smoking, drinking heavily, being overweight, or not exercising. Even behavioral changes are not highly correlated; the largest correlation in the table is less than .1. The investment model is thus incomplete along this dimension as well.

What, then, are the alternatives? There are three possibilities: health behaviors differ because of differing information; because of genetic differences; and because of situational factors that are specific to the person and behavior. In the remainder of the paper, we examine these different factors.

I.   Framework

To understand our analysis, we begin with a simple two-period model of individual choices, where individual i chooses health related actions (denoted a_j for action j) to maximize expected discounted lifetime utility:

$$W = \sum_j (b_j^i a_j - .5a_j^{\,2}) + \left(\bar{\pi} - \sum_j \tilde{\pi}_j^i a_j \right) \beta_i V_i, \qquad (1)$$

where b_j^i reflects the individual-specific current return to the action (net of price), $\left(\bar{\pi} - \sum_j \tilde{\pi}_j^i a_j \right)$ is the probability of survival to the next period, and $\tilde{\pi}_j^i$ is the self-perceived effect of action j on the survival probability. β_i is the individual-specific discount rate, and V_i is the value to surviving into the next period. The first order condition for action j is: $a_j = b_j^i - \tilde{\pi}_j^i \beta_i V_i$.

We assume that $b_j^i = \bar{b}_i + \bar{b}_j + \varepsilon_j^i$, where $\bar{b}_i$ is the individual tendency to engage in these activities, $\bar{b}_j$ is the average preference for this activity in society, and ε_j^i is an activity and person-specific idiosyncratic taste. We normalize $\bar{b}_i$ and ε_j^i to mean zero and assume they are independent of each other and $\bar{b}_j$. Similarly, $\tilde{\pi}_j^i = \bar{\pi}_j + \mu_j^i$ where $\bar{\pi}_j$ represents the society-wide average belief (which may or may not be correct) and μ_j^i is an idiosyncratic individual factor with mean zero and no correlation with the other variables. $\overline{\beta V}$ is the expectation of $\beta_i V_i$.

The variance of the action across individuals is given by:

$$Var(a_j) = \overline{\pi}_j^2 Var(\beta_i V_i) + \left(Var(\beta_i V_i) + \overline{\beta V}^2\right)Var(\mu_j^i) + Var(\overline{b}_i + \varepsilon_j^i) \qquad (2)$$

The first term is the heterogeneity related to discount rates or the value of life. The second term relates to the heterogeneity of beliefs about the impact of action j: when people view the behavior in question differentially, this will influence the observed variability in outcomes. The third term reflects heterogeneity in preferences, some of which are common across behaviors and others of which are idiosyncratic.

The correlation across people between observed actions j and k is

$$corr(a_j, a_k) = (\overline{\pi}_j \overline{\pi}_k Var(\beta_i V_i) + var(\overline{b}_i)) \Big/ stddev(a_j)stddev(a_k).$$ The numerator consists of variability in standard investment effects (the first term) and variability in individual preferences (the second term).

The small correlation we observe between different behaviors implies that neither variability in standard investment factors nor in cross-behavior personal preferences is that large – the two together do not top 10 percent of total variability. Understanding the heterogeneity in behaviors requires us instead to understand the variability in perceived health consequences, $\overline{\beta V}^2 Var(\mu_j^i)$, and tastes, $Var(\varepsilon_j^i)$.

Since many of our health outcomes are discrete, rather than continuous, we extend the model to address the continuous outcomes. We assume that there is remains a latent variable a_j, and to simplify the discussion, we assume that this latent variable equals $\varepsilon_j^i - \overline{\pi}_j \beta_i V_i$, the idiosyncratic taste for the variable minus the expected health costs. Individuals undertake the discrete action if and only if a_j is positive. Our goal is to assess $\overline{\pi}_j^2 Var(\beta_i V_i) / \overline{\pi}_j^2 \left(Var(\beta_i V_i) + Var(\varepsilon_j^i)\right)$, that share of the variance that is accounted for by heterogeneity in investment factors.

To make progress, some assumption about the distribution of the factors is necessary. For simplicity, we assume that both variables are uniform, and we normalize so that $\beta_i V_i$ is uniformly distributed across the population on the interval [0, 1]. The variable ε_j^i is assumed to be uniformly distributed on the interval $\left[\theta_j - .5e_j, \theta_j + .5e_j\right]$.

Straightforward algebra shows that $\overline{\pi}_j^2 Var(\beta_i V_i) / \overline{\pi}_j^2\left(Var(\beta_i V_i) + Var(\varepsilon_j^i)\right) = \dfrac{\overline{\pi}_j^2}{\overline{\pi}_j^2 + e_j^2}$.

For discrete outcome k, we similarly assume $a_k = \varepsilon_k^i - \overline{\pi}_k \beta_i V_i$, where ε_k^i is distributed uniformly on the interval $\left[\theta_k - .5e_k, \theta_k + .5e_k\right]$. Again, the action is undertaken if and only if a_k is positive. Finally, we assume the $e_j > 2Max\left(\theta_j, \overline{\pi}_j - \theta_j\right)$ and $e_k > 2Max\left(\theta_k, \overline{\pi}_k - \theta_k\right)$ so that for any value of $\beta_i V_i$, it is possible to have both outcomes.

Denoting σ_{jk} as the covariance between the two discrete outcomes, and p_j and p_k as the probability of the two outcomes, it follows that

$$\frac{\left|\sigma_{jk}\right|}{\dfrac{1}{12\sqrt{p_j(1-p_j)p_k(1-p_k)}} + \left|\sigma_{jk}\right|} = \frac{\left|\overline{\pi}_j \overline{\pi}_k\right|}{e_j e_k + \left|\overline{\pi}_j \overline{\pi}_k\right|} \geq \frac{\left|\overline{\pi}_j \overline{\pi}_k\right|}{\sqrt{\overline{\pi}_j^2 + e_j^2}\sqrt{\overline{\pi}_k^2 + e_k^2}}.$$ The term

$$\frac{\left|\sigma_{jk}\right|}{\dfrac{1}{12\sqrt{p_j(1-p_j)p_k(1-p_k)}} + \left|\sigma_{jk}\right|}$$ therefore provides us with an upper bound on the

geometric average of the value of $\dfrac{\overline{\pi}_j^2}{\overline{\pi}_j^2 + e_j^2}$ and $\dfrac{\overline{\pi}_k^2}{\overline{\pi}_k^2 + e_k^2}$, the share of variance in each

outcome accounted for by the investment fundamentals.

Using this formula, we have adjusted Table 1 so it contains both the estimated correlation coefficients and the value of $\dfrac{\left|\overline{\pi}_j\,\overline{\pi}_k\right|}{e_j e_k + \left|\overline{\pi}_j\,\overline{\pi}_k\right|}$ implied by the correlation coefficient, the latter in [.]s below the correlation. In all cases, the procedure increases the estimated correlation coefficients. The estimates of $\dfrac{\left|\overline{\pi}_j\,\overline{\pi}_k\right|}{e_j e_k + \left|\overline{\pi}_j\,\overline{\pi}_k\right|}$ cluster around .15. Again, these figures are low and suggest that the bulk of heterogeneity in behaviors cannot be explained by variation in the common investment-related characteristics.

II.   Health Beliefs and Health Behaviors

To examine the importance of differing health beliefs, we relate the behaviors described above to survey response about what people believe are the consequences of these actions. The 1990 NHIS asks a number of questions about the potential adverse effects of different behaviors. A typical question is of the form: "Does cigarette smoking increase the chance of developing lung cancer?," with possible responses of: definitely increases; probably increases; probably does not increase; and definitely does not increase. We demarcate people by whether they believe the behavior in question definitely increases the risk or not.[2] For smoking, questions are asked about the link to cardiovascular disease, cancer, and respiratory problems. Questions are also asked about the harmful effects of alcohol, being overweight, and having high blood pressure, but not about failure to receive a mammogram.

In each case, the vast majority of people believe that these behaviors are harmful. For example, 74 percent of people respond that cigarette smoking is definitely related to cardiovascular disease, 86 percent believe it is definitely related to cancer, and 85 percent believe it is definitely related to respiratory problems. For alcohol use, being overweight, and having hypertension, the corresponding figures are 87 percent, 75 percent, and 69 percent. Beliefs are also generally related to demographics, as we would expect. Better educated and higher income people are more likely to believe that most of the behaviors are harmful to health. Racial effects are not very big.

That is not to say that the belief measures are without problems. Clearly, they measure true belief with error. More importantly, these beliefs may rationalize, rather than cause behavior. But rationalization would tend to overstate the importance of beliefs.

Table 3 shows the relation between beliefs, socioeconomic factors, and health behaviors.[3] For each of the behaviors described above, the first rows report the effect of education and income on engaging in the behavior (also included in the regressions, but not reported, are five year age-sex groups, dummy variables for blacks, other races, and Hispanics, and a dummy variable for being married).

People who invest more in education generally invest more in health as well; there is a significant education effect in most regressions. Indeed, the education effect is large; behaviors differ by as much as 10 percentage points between college graduates and high school dropouts. Income is also generally associated with better health. However, the explanatory power of these socioeconomic factors is low. The R^2's from socioeconomic factors alone are reported in the bottom of the table[4] and range from 2 percent for using hypertension medication to 9 percent for receiving a mammogram on

the recommended schedule. These results confirm the low correlations noted in table 1. Once again, these results are not unique to the data or specific behaviors indicated. Appendix table 2 shows that the same low R^2 for socioeconomic factors is found in the BRFSS data as well.

The next rows report the coefficients on beliefs. Beliefs are almost always correlated with behavior in the expected way. People who think that smoking or drinking poses a greater risk are unsurprisingly less likely to smoke or drink. The magnitude of these effects ranges up to 15 percentage points for smoking, but is much weaker for the other variables. But again, the ability of these variables to explain individual variation is small. Including belief variables in the regressions increases the R^2 by only 5 percent in the case of smoking, and even less for the other behaviors. Even recognizing that beliefs are measured with error, the very low explanatory power and possibility for *ex post* rationalization makes it hard to accept that belief heterogeneity explains differences in individual behaviors.

IV.   The Individual vs. the Situation

There are two primary rivals to the economic view of health investments: the genetic view that human heterogeneity is driven primarily by genetic differences, and the psychological view that human choices are driven primarily by randomly-encountered situational differences – a chance encounter with a particular peer group, perhaps, or a random health shock that affects that individual and not others.

Our first attempt to differentiate between these behaviors is to estimate the importance of fixed individual effects in different forms of health behaviors. This is an

upper bound on the importance of genetics or other behavior-specific factors. To do this, we use data from the Health and Retirement Study (HRS). We gauge the importance of individual effects using a random effects regression model, controlling in the model for aggregate year effects.[5]

The first column of Table 4 lists the partial-R^2 of the individual effects. The R^2 ranges from 38 percent for vigorous exercise, to nearly 90 percent for body mass index. Smoking shows a high degree of permanence (79 percent R^2) and drinking less (51 percent R^2). The high R^2's for smoking and obesity can be explained by many different theories. Genetics might be important, but so too might be addiction (rational or myopic) or peer effects with relatively stable peer groups.

We consider first the potential importance of neighborhoods. Neighborhoods might matter for health behaviors because of peer effects, or because of correlated shocks such as high or low tax rates on cigarettes. To gauge the importance of neighborhood effects, we have estimated the added R^2 in table 3 that comes from including geographic controls at the Primary Sampling Unit (roughly MSA) level. Addition of the geographic controls raises the R^2 by two to seven percent.[6] The relatively modest explanatory power of area effects suggests that these gross measures of situation are not where the bulk of the situational variation results from.

We can say more about the impact of genetics on health behaviors. The Minnesota Twin Registry[7] has data on many health behaviors among all twins born in Minnesota between 1936 and 1955, nearly 4,000 individuals in total. Following the standard literature on heritability, we compare the correlation of behavior between

monozygotic (identical) and dizygotic (fraternal) twins. The genetic contribution to any behavior is twice the difference between the correlation in these groups.

The second column of Table 4 reports the genetic contribution to variability in health behaviors. The highest heritability estimate is for body mass index – 72 percent of the variation across people in body mass is estimated to be the result of purely genetic factors. The heritability for other variables is smaller: 26 percent for exercise, 30 percent for smoking and 38 percent for heavy drinking. Comparing columns 1 and 2 of table 4, about one-half of the individual variation in these behaviors results from a genetic predisposition to engage in them.

Our bottom line estimate is that variation in health behaviors results from two primary factors: genetics, and behavior-specific situational influences. We have not to date identified what the relevant situational influences are, but that is clearly a high priority for future research.

References

McGinness, J. Michael, and William H. Foege, "Actual Causes of Death in the United States," *Journal of the American Medical Association*, 1993, 270(18), pp. 2207-12.

Grossman, Michael, On the Concept of Health Capital and the Demand for Health," *The Journal of Political Economy*, 1972, 80(2), Mar.-Apr., pp. 223-55.

Lykken, David T., Thomas J. Bouchard, Matthew McGue, and Auke Tellegen, "The Minnesota Twin Family Registry: Some Initial Findings," *Acta Geneticae Medicae et Gemellologiae*, 1990, 39, 35-70.

Table 1: The Correlation of Health Behaviors

	Smoker	Heavy Drinker	Obese	Use HTN Medication	Mammogram as recommended
Mean	21%	8%	17%	68%	53%
N	18,076	18,076	17,817	5,850	10,877
Smoker	1.000				
Drinker	.160	1.000			
	[.176]				
Obese	.099	.003	1.000		
	[.154]	[.004]			
Use HTN medication	-.086	-.073	.018	1.000	
	[.164]	[.106]	[.036]		
Mammogram as recommended	-.073	.001	-.050	.063	1.000
	[.151]	[.003]	[.101]	[.149]	

Note: The data are from the 1990 National Health Interview Survey. The sample is the population over age 45, restricted to people prescribed anti-hypertensive medication for the medication use variable and women for the mammography variable. All data are self-reports. The numbers is brackets refer to the estimates of the ratio of variance of common preferences to the total variance of preferences based on the assumptions given in the text.

	Quit Smoking	Stop Heavy Drinking	Start Vigorous Exercise	Lose Weight
Table 2: The Correlation of Changes in Health Behaviors

	Quit Smoking	Stop Heavy Drinking	Start Vigorous Exercise	Lose Weight
Mean	15%	44%	27%	6%
Quit smoking	1.000			
Stop heavy drinking	.090	1.000		
Start vigorous exercise	-.009	.053	1.000	
Lose weight	-.012	.069	.036	1.000

Note: The data are from the Health and Retirement Study. Smoking and BMI are available for six waves of the Health and Retirement Study; heavy drinking and vigorous exercise are available for four waves.

Table 3: Socioeconomics, Beliefs, and Geography

Independent Variable	Smoker	Heavy Drinker	Obese	Use HTN Medication	Mammogram as recommended
Education					
<High school	.019**	-.011**	.059**	-.026*	-.093**
	(.008)	(.005)	(.008)	(.014)	(.012)
Some college	-.021**	-.021**	-.005	.016	.056**
	(.009)	(.006)	(.009)	(.017)	(.014)
College +	-.068**	-.047**	-.037**	.004	.001
	(.009)	(.006)	(.009)	(.018)	(.015)
ln(Income)	-.035**	.007*	-.023**	.007	.074**
	(.006)	(.004)	(.005)	(.010)	(.008)
Beliefs					
1st dimension	-.099**	-.023**	-.032**	.022*	---
	(.008)	(.006)	(.007)	(.012)	
2nd dimension	-.155**	---	---	---	---
	(.011)				
3rd dimension	-.088**	---	---	---	---
	(.011)				
N	16,189	17,073	16,884	5,373	10,877
Share of variance accounted for by:					
Demographics	.065	.061	.030	.023	.091
Beliefs	.051	.001	.001	.002	---
Geography	.024	.025	.019	.074	.047
Total R^2	.140	.086	.050	.089	.138

Note: All regressions are linear probability models with controls for five year age and sex groups; racial and ethnic dummy variables; a dummy variable for being married; dummies for missing education and income; and area dummy variables. Income is asked in buckets; people are assigned the median income of their bucket. * (**) indicates statistical significance at the 10 percent (5 percent) level.

Table 4: Genetics and Individual Variability in Health Behaviors

Behavior	HRS Data Share of Variance Accounted for by Individual Effects	Minnesota Twins Data Share of Variance Accounted for by Genetics
Smoking	79%	30%
Heavy drinking	51%	38%
Vigorous exercise	38%	26%
BMI	88%	72%

Note: The share of variance accounted for by individual effects is the R^2 on the individual random effects, taking out the variance attributable to time. Smoking and BMI are available for six waves of the HRS; heavy drinking and vigorous exercise are available for four waves.

Appendix Table 1: Correlation of Health Behaviors, BRFSS

	Smoker	Drinker	Obese	Use seatbelt	Flu shot	Colorectal screen	Mammo-gram	Pap Smear	PSA test
Mean	18%	9%	26%	83%	66%	44%	77%	75%	75%
N	136,189	136,629	130,553	134,759	51,082	110,339	68,041	82,450	40,704
Smoker	1.000								
Drinker	.129	1.000							
Obese	.025	.001	1.000						
Use seatbelt	-.071	-.061	-.083	1.000					
Flu shot	-.092	-.051	-.025	.063	1.000				
Colorectal screen	-.109	-.036	-.009	.056	.176	1.000			
Mammogram	-.108	-.008	-.014	.089	.190	.215	1.000		
Pap smear	-.055	.010	-.013	.060	.101	.121	.469	1.000	
PSA test	-.145	-.064	.012	.059	.188	.320	---	---	1.000

Note: Data are from the 2002 Behavioral Risk Factor Surveillance System. Samples include everyone 45 and older with the exception of flu shots (ages 65+), and colorectal screen, mammograms, and PSA tests (ages 50+). Mammograms and Pap smears are for women only. PSA tests are for men only.

Appendix Table 2: Socioeconomics, Geography, and Health Behaviors, BRFSS

Independent Variable	Smoker	Drinker	Obese	Use seatbelt	Flu shot	Colorectal screen	Mammo- gram	Pap Smear	PSA test
Education									
<High school	.030**	-.007**	.038**	-.023**	-.039**	-.046**	-.061**	-.091**	-.071**
	(.003)	(.003)	(.004)	(.003)	(.006)	(.005)	(.005)	(.011)	(.007)
Some college	-.020**	-.010**	-.008**	.018**	.025**	.038**	.011**	.017**	.048**
	(.003)	(.002)	(.003)	(.003)	(.006)	(.004)	(.004)	(.004)	(.006)
College +	-.104**	-.034**	-.063**	.044**	.026**	.084**	.031**	.040**	.072**
	(.003)	(.002)	(.003)	(.003)	(.006)	(.004)	(.004)	(.004)	(.006)
ln(Income)	-.033**	.017**	-.032**	.000	.014**	.043**	.048**	.041**	.064**
	(.002)	(.001)	(.002)	(.002)	(.004)	(.003)	(.003)	(.003)	(.004)
N	135,393	135,818	129,844	133,970	50,650	109,623	67,608	82,068	40,451
Share of variance accounted for by:									
Demographics	.078	.064	.037	.028	.041	.060	.047	.081	.095
Geography	.014	.015	.017	.049	.045	.022	.069	.060	.035
Total R^2	.092	.079	.054	.077	.086	.082	.116	.141	.130

Note: Data are from the 2002 Behavioral Risk Factor Surveillance System. Samples include everyone 45 and older with the exception of flu shots (ages 65+), and colorectal screen, mammograms, and PSA tests (ages 50+). Mammograms and Pap smears are for women only. PSA tests are for men only. ** denotes statistical significance at the 5 percent level.

[1] One might worry that smoking is metabolically associated with lower weight. We have experimented with adding up to 10 pounds to the weight of smokers, a rough estimate of weight gain after smoking cessation. That does not materially affect the correlations reported in Table 1.

[2] Virtually everyone believes these behaviors are at least somewhat related to poor outcomes. In many cases, multiple questions are asked. We denote a person as believing the activity is definitely bad if they report it definitely causes at least one outcome.

[3] Area-level dummy variables are also included in the regressions, as discussed below.

[4] The incremental R^2's are calculated sequentially. The R^2 for socioeconomics is based on a regression with those variables only; the R^2 for beliefs is the incremental explanatory power from adding beliefs to the model with SES, and similarly for geography.

[5] We use random effects rather than fixed effects because of the short panel.

[6] A similar finding is true in the BRFSS data.

[7] These data were collected by Jere R. Behrman, Mark R. Rosenzweig, and Paul Taubman of the Economics Department at the University of Pennsylvania with funds from the National Center on the Educational Quality of the Workforce, the Economics Institute Research Fund, the Boettner Research Fund, the Population Study Center NIA Supplement, and the University Research Foundation -- all of the University of Pennsylvania. The survey instrument was designed in collaboration with the Temple University Institute of Survey Research. The data were collected, under the leadership of David T. Lykken, then Director of the Minnesota Twin/Family Registry (MTR), by the staff of the MTR. The data are described in David T. Lykken, et al., (1990).