

Does Education Help with Complex Health Decisions: Evidence from Cancer Screening.

Fabian Lange
(Yale University and IZA)

January 19h, 2010

Abstract

This paper uses data on real and perceived cancer risks and cancer screening behavior to test whether education helps to make better informed health decisions. I propose that if educated individuals are better informed, then they will be more likely to incorporate variation in risk factors when they report their personal cancer risk. As risk varies, they will also react more strongly by adopting preventive behaviors such as cancer screening. I find support for both predictions. Using data on attitudes surrounding breast health, I explore a possible mechanism: educated women are more receptive to scientific evidence and hold fewer non-scientific beliefs.

Keywords: Education, Allocative Efficiency, Health

JEL Codes: I12, I10, I20, D83

Acknowledgement 1 *Correspondence please to [fabian.lange\[at\]yale.edu](mailto:fabian.lange[at]yale.edu). I owe thanks to Jason Fletcher and Doug McKee for extremely helpful feedback on a late draft of this paper as well as to Keith Chen for many discussions throughout the project. I also would like to thank seminar participants at the University of Illinois - Chicago, Harvard University, Northwestern University, Suny New York and Cuny New York. Mistakes and omissions are obviously my fault.*

1 Introduction:

One of the undisputed facts in health economics is that education and health are very highly correlated, so much so that Grossman and Kaestner (1997) refer to education as the most important socioeconomic correlate of good health. A number of explanations have been offered for this tight correlation. Education is thought to raise the demand for health, both because health is a normal good and because individuals have an incentive to invest in health to protect their education investments. Causality might run the other way, because longer life-expectancies raise the present value of financial returns to education and because health can directly affect the ability to learn. And, it is also possible that both health and education investments are caused by variation in preference parameters such as the discount rate. Education might also allow agents to produce health more efficiently, possibly because education enables agents to better process and act on health information. This hypothesis is known as the allocative efficiency hypothesis.

Understanding whether education increases the ability to process health information is becoming increasingly important as medical progress accelerates. Modern medicine can treat or manage an increasing number of diseases. Yet, modern medical care often requires patients to adhere to complex treatment schedules and patients face a bewildering array of options. If the skills required to negotiate the modern health care system are partially acquired through education, then lack of education will increasingly constrain patients' ability to participate in the gains from medical progress.

The interaction between the lack of skills relevant for health choices and continuing medical progress might therefore partially explain why health inequalities are increasing rapidly (Meara, Richards, and Cutler (2008) and Cutler, Lange, Meara, Richards, and Ruhm (2010)). If policy makers want to expand the benefits of modern medical care to the less educated, they need to understand whether the less educated suffer from worse health because they can't afford health care, or because they lack the relevant health-care knowledge and decision making skills. One of the core tasks of empirical research in health economics is therefore to test whether education increases

the efficiency of health production and specifically whether it increases allocative efficiency.¹

Empirical research into the allocative efficiency hypothesis has so far been limited by the fact that few data-sets contain evidence on health knowledge and health investments at the same time. To our knowledge, Kenkel (1991a, 1991b) is the only study that simultaneously examines both health behaviors and health knowledge. He showed that the educated are more knowledgeable about health risks from smoking, drinking, and lack of exercise and also showed that there is a positive correlations between knowledge about health risks and healthy life-style choices. In part due to the fact that few surveys contain data on both information and health behaviors, the evidence on allocative efficiency of education in making health investments is therefore still sparse.

In this paper, we use data from the National Health Interview Survey (NHIS) on cancer screening behavior and from subjective statements about cancer risks to provide new evidence on the allocative efficiency hypothesis.² In particular, we examine whether the educated are better informed about their personal risk for developing cancer and whether they are more likely to act on this knowledge. We use a simple model to derive a proposition about how screening and risk assessments vary across education if (i) the educated are better informed or alternatively (ii) the educated are willing to expend more resource on reducing mortality risks.

We present evidence from both screening and from risk assessments that support the allocative efficiency hypothesis. The finding based on subjective risks is particularly interesting, because it can not be explained by variation in the demand for health across education. The evidence is consistent across various cancers: educated individuals are more likely to be screened for can-

¹In Section 2, I review in more detail the literature that tests for whether education affects the production of health directly.

²A number of researchers have directly examined the role of education in preventive care. In particular, Fletcher and Friswold (2009) as well as Cutler and Lleras-Muney (2007, 2008) show that education is associated with an increase in the probability of preventive care, including various cancer screens. Kenkel (1994) argues that preventive care has primarily a investment character as opposed to the consumptive character of many other health inputs.

cer if they are objectively (based on current medical evidence) of higher risk of developing cancer. And, similarly, educated individuals are more likely to incorporate the presence of objective risk factors when assessing their personal cancer risk.

We then briefly consider survey evidence from an alternative source that sheds some light on why the less educated seem to be less aware of the risks they are facing. This evidence suggests that the less educated are generally more sceptical about science-based medicine and more likely to believe in non-scientific explanations for why individuals develop cancer. It is plausible that scepticism or downright hostility towards science based medicine contributes to less efficient health care decision making among the less educated.

We begin in Section 2 with a review of the literature that tests for a direct role of education in producing health, possibly due to allocative efficiency. In Section 3, we present a simple stylized model of health decision making that organizes our empirical analysis. Section 4 presents the data. Section 5 contains the main empirical results based on the NHIS. Section 6 discuss the evidence on attitudes towards science-based medicine. Section 7 concludes.

2 Background

In his path-breaking analysis of health as the outcome of an investment process, Grossman (1972) from the start proposed that education makes individuals more productive in producing health. This hypothesis is known as the productive efficiency hypothesis. In the original paper and in subsequent contributions, he and followers showed how one can use data on inputs and input prices as well as health outcomes to test for productive efficiency of education. We begin by discussing two papers that we find particularly informative: Gilleskie and Harrison (1998) and Wagstaff (1986).³ These papers provide some support for the productive efficiency hypothesis, but they also illustrate how difficult it is to assemble all the data required to test

³For a more comprehensive review of the literature see Grossman and Kaestner (1997) and more recently Grossman (2005).

for productive efficiency. There are many inputs that are potentially relevant and are very difficult to observe. Unobserved inputs are often likely to correlate with education, making it difficult to causally interpret observed partial correlations between education and health.⁴ Therefore, a number of authors have proposed auxiliary assumptions in specific health care settings to derive testable implications of the productive efficiency hypothesis. Examples include Goldman and Smith’s (2002) study of diabetes and HIV self-management and Lichtenberg and Lleras-Muney’s (2002) analysis of how education affects the adoption of new medical technologies. Their contribution supports the allocative efficiency hypothesis for which Kenkel (1991a, 1991b) provided survey evidence on smoking, alcohol use, and exercise. Our work contributes to this evidence in support of the allocative efficiency hypothesis using new data and with a new methodological approach.

2.1 Production Functions and Health Demand

To organize the evidence in Gilleskie and Harrison (1998) and Wagstaff (1986), we consider a static version of Grossman’s (1972) dynamic model of health. At the heart of the model is a production function that maps health endowments Ω_i , market inputs M_i , non-market inputs B_i , and time T_i into produced health. Health endowments differ from inputs in that they are not subject to agents control. Market inputs M_i are such inputs as doctor visits and prescription drugs. Non-market inputs include choices such as exercise, smoking, and nutrition but also the effort and pain that are involved in medical treatment choices. Non-market inputs are distinct from market inputs in that the former enter both in the utility function and the budget set, while the latter enter only through the budget set. Empirically, a major distinction is that the shadow prices of inputs B_i are typically much harder to measure than the prices associated with M_i .

⁴The problem is compounded by the fact that health, the output of the production process we strive to analyze, is an abstract concept with no clear-cut empirical counterpart. Various health measures can be analyzed and all are likely important, but the weights consumers place on these various aspects of health are not directly observed and might differ from how the empirical analyst weights health.

The production function (1) maps $(T_i, M_i, B_i, \Omega_i)$ into health:

$$H_i = F(T_i, M_i, B_i, \Omega_i; S_i) \quad (1)$$

Schooling S_i enters directly into this production function. The productive efficiency hypothesis is that $\frac{\partial F}{\partial S_i} > 0$. The challenge is to test $\frac{\partial F}{\partial S_i} > 0$ given that schooling is potential endogenous to health inputs and endowments.

The choices of consumers are ordered by a utility function $U(x_i, B_i, H_i; S_i)$ over consumption x_i , non-market inputs B_i and health H_i . Health enters the utility function and there is therefore a consumption motive for expending resources on health production. The utility function also depends on schooling directly, allowing for the possibility that schooling correlates with preference parameters (Fuchs (1982)). Individuals choose health inputs to maximize the utility function subject to a budget set $G(H_i, T_i, M_i, B_i, x_i; S_i, W_i, p_M, p_B) \geq 0$. Health enters directly into the budget set because it might raise individuals productivity. Schooling affects this budget set primarily through the wage, but also potentially through the timing of earnings. Maximizing preferences subject to the budget set will deliver demand functions for inputs as functions of schooling, health endowments, as well as wages and input prices.

2.2 Testing for Productive Efficiency with Rich Data on Health, Inputs, and Input Prices.

The most straightforward approach to test the productive efficiency hypothesis relies on estimating the production function (1) directly. Assuming that we observe the inputs (M_i, B_i, T_i) , the health endowment Ω_i as well as data on H_i and schooling S_i , then we can immediately estimate the production function and can test $\frac{\partial F(T, M, B, \Omega; S)}{\partial S}$ directly.

Conceptually, the advantage of this approach is that we do not have to model or estimate the medical care decision itself. With sufficiently high quality data on health endowments, inputs and outputs, we can estimate the production function (1) directly and do not require any additional structure

such as that incorporated in specific utility functions, the budget set or even the assumption that individuals maximize utility. In practice, the production function approach will often fail because we typically only observe subsets of the relevant health inputs, endowments, and health outcomes.

Gilleskie and Harrison (1998) implement this approach to testing for productive efficiency using self-reported health status as their outcome.⁵ Their input measures are counts of both curative and preventive medical care visits. They measure behavioral inputs using smoking and physical exercise and proxy for health status using the number of chronic conditions and the body mass index of individuals. They find evidence that education affects health positively even after controlling for the vector of inputs and measures of health endowments available to them. How plausible one finds this evidence depends on the willingness to assume that schooling is exogenous to unmeasured inputs and health endowments.

Clearly, this approach requires very good data on medical and behavioral inputs, the endowments of health, and the health outcomes. It is possible to substitute for data on observed inputs with price data if one is willing to assume utility maximization. The consumer problem results in derived demand functions for medical inputs $M^D(p_M, W, \tilde{p}_B, H, S)$ where \tilde{p}_B is the full (shadow) price of behavioral modifications which consists of both the financial cost p_B and of the dollar-denominated utility value of B. To simplify the discussion, we assume that behavioral modifications are secondary in the production of health and write $M^D(p_M, W, H, S)$.

Under the productive efficiency hypothesis, educated individuals require fewer inputs to produce a given amount of health. If education is factor neutral in equation (1), if inputs are normal, and if we observe all input prices, then we can test the productive efficiency hypothesis with data on a single health input. In particular, we can test whether the more educated

⁵We do not have the room here to do the analysis by Gilleskie and Harrison (1998) full justice. Gilleskie and Harrison also estimate demand functions for medical inputs and they estimate dynamic specifications of health production, as opposed to the static specifications we consider here. Finally, Gilleskie and Harrison find evidence that part of the increase of health associated with education is due to a higher consumption of medical care among the educated.

consume less of this input conditional on health outcomes and input prices.

Wagstaff (1986) applies this approach on Danish data, using doctor visits, hospital stays, and the consumption of prescription drugs as input measures. Given the generosity of the Danish health care system, Wagstaff argues that patients in his data face the same (zero) financial costs for all medical inputs. He controls for time costs using the wage and the distance to the next doctor or hospital. Wagstaff controls for the stock of health using a principle component extracted from multiple health indicators. Interestingly, Wagstaff finds that $\frac{\partial M^D(p_M, W, H, S)}{\partial S} > 0$ which contradicts the productive efficiency hypothesis: it seems that the educated require more medical inputs to achieve a given level of health.

Wagstaff also estimates a demand function for the stock of health. We can write this demand for health as $H^D(\pi(p_M, w; S), w, S)$ where $\pi(\cdot)$ is the unit cost of health. To construct this unit cost of health requires fully specifying the health production function. An alternative approach is to let all prices (p_M, p_B, w) enter in a flexible form. If we are willing to assume that there is no consumption motive of education, then health demanded is $H(\pi(p_M, p_B, w; S), w)$ and we can test the productive efficiency hypothesis by testing $\frac{\partial H}{\partial S} = \frac{\partial H}{\partial \pi} \frac{\partial \pi}{\partial S} > 0$. As long as the demand for health declines in its price, we can interpret a positive sign of $\frac{\partial H}{\partial S}$ as indicating that the unit cost of producing health is declining in schooling. Wagstaff indeed finds that $\frac{\partial H}{\partial S}$ is positive, a finding that is consistent with the productive efficiency motive.

Wagstaff (1986) therefore reports conflicting evidence depending on whether he estimates $M(p_M, W, H, S)$ or $H(p_M, p_B, w, S)$. These conflicting findings are typical for the literature. They might reflect differences in the set of maintained auxiliary assumptions that are required to use either $M(p_M, W, H, S)$ or $H(p_M, p_B, w, S)$ to test for productive efficiency. For instance, if there is a consumption motive to the demand for health, then it is possible that both $\frac{\partial H}{\partial S}$ and $\frac{\partial M}{\partial S}$ are positive. Alternatively, these conflicting findings might be due to missing data on inputs that are endogenous to education.

2.3 Alternative Approaches to Testing for Productive and Allocative Efficiency

The extraordinarily detailed data on health inputs, prices, and outputs required to implement the above described testing procedures is rarely available. Therefore, a number of researchers instead exploit auxiliary assumptions in specific health care settings to generate additional predictions that can be tested in the available data.

Goldman and Smith (2002) represents a particularly instructive example.⁶ They examine the role of patient self-management in generating differences in health outcomes by education. Using data from HIV and diabetes patients, they find that educated patients are indeed better at adhering to complex treatment regimens and that this has positive effects of the health of these patients. This finding represents direct evidence for productive efficiency if self-management does not represent an input into the production of health in the traditional sense, that is if it does not impose costs. If self-management is however costly, then it is not clear whether we observe higher adherence among the educated because they are willing to expend more resources on health or because they are better at producing health. In our opinion, Goldman and Smith's evidence from diabetes care is more convincing than that related to adherence to the HIV treatment regime (called Highly Active AntiRetroviral Therapy = HAART). In their data, an individual is coded as adhering to the HAART regime if she takes the prescribed drugs in all of the last 7 days prior to the survey. These drugs are costly, and therefore the adherence to HAART reflects partially the costs associated with the treatment regime. In the case of diabetes, Goldman and Smith examine whether patients have been consistent in their treatment regime or

⁶Related is the work by Goldman and Lakdawalla (2001), who examine how health inequality varies across various cardiovascular disease and HIV. They find that technological progress differentially affects inequality in these diseases, a finding that they rationalize with the particular form technological progress has taken in these diseases. Clearly, it is possible to rationalize any variation in health inequality with technological progress by assuming this technological progress to be either augmenting or substituting for education. However, if we accept the claims of Goldman and Lakdawalla (2001) with respect to the skill-bias of technological progress for the various disease considered in their study, then this work represents additional evidence in favor of the productive efficiency hypothesis.

whether they have switched repeatedly among different treatment options. Overall, the evidence from Goldman and Smith supports the productive efficiency hypothesis.

Lichtenberg and Lleras-Muney (2002) follow a different approach. They argue that the educated are better placed to benefit from technological progress in the medical sciences. In support, they show that the educated are disproportionately more likely to use newly approved drugs, even controlling for income and insurance status. This finding might explain the recent widening in the education gradient in mortality (Cutler, Lange, Meara, Richards, and Ruhm (2010)).

Kenkel (1991a, 1991b) provides survey evidence in favor of the allocative efficiency hypothesis, showing that the educated are better informed about health risks associated with smoking, drinking, and lack of exercise. Knowledge about these risks correlates with healthier choices, thus confirming the allocative efficiency hypothesis. Our work below is related in that we examine how individual knowledge of personal cancer risks correlates with cancer screening behavior and we do find support for the allocative efficiency hypothesis from a variety of different screening behaviors.

3 A Model of Cancer Screening and Subjective Risk

In this Section, we develop a simple stylized model of health demand and information. This stylized model implies that better informed individuals react more strongly to the presence of objective risk factors in their screening decisions, but also in their subjective assessments of how much risk they face. The prediction on screening forms the basis of our first test of the allocative efficiency hypothesis and this prediction is indeed borne out in the data. However, simple demand differences across education can (but need not) generate the same prediction for screening behavior. It is for this reason that the data on subjective risk is important. It allows us to derive and test an implication of allocative efficiency using the subjective risk data that is

not implied by variation in the willingness-to-pay across education.

Let individuals belong to either a low and high risk group of developing cancer $d_i \in \{L, H\}$. These groups differ in the probability of developing cancer μ_d and $\mu_H > \mu_L$. Empirically, we will associate these risk classes with the objective risk factors that are identified in the medical literature, the most important of which is the presence of cancer in the family. The information problem that agents face is to predict whether or not they belong to the high or low risk group.

Let $cr_i \in \{0, 1\}$ indicate the presence of cancer and let θ_i indicate screening.⁷ Cancer screening generates survival gains g among those with cancer because of early diagnosis

$$g = \Pr(\text{survival} | cr_i = 1, \theta_i = 1) - \Pr(\text{survival} | cr_i = 1, \theta_i = 0)$$

The marginal rate of substitution of consumption for survival is known as the Value of a Statistical Life. Then, if we denote the costs of screening by c_i with distribution function $F(\cdot)$, we can represent the screening decision as:

$$E[\mu | \rho_i] g VSL_i - c_i \geq 0 \tag{2}$$

The signal ρ_i introduces the idea that information about individual cancer risks might differ across individuals. Individuals receive a signal $\rho_i \in \{0, 1\}$ that reveals information about their risk class. Assume that receiving $\rho_i = 1$ perfectly reveals that $d_i = H$:

$$\begin{aligned} \Pr(\rho_i = 1 | d_i = L) &= 0 \\ \Pr(\rho_i = 1 | d_i = H) &= \pi_\rho \in (0, 1) \end{aligned} \tag{3}$$

The parameter π_ρ determines how informed individuals are in the sense that if π_ρ increases, then individuals are more likely to learn that they are of high risk if they indeed belong to the high risk group. This simple model has implications about how subjective cancer risk, screening θ_i and objective

⁷Note: $\Pr(cr_i | I) = E[\mu | I]$ where I is an arbitrary information set.

risk d_i are related to the quality of information π_ρ available to agents as well as the willingness-to-pay for survival gains VSL_i . We summarize these implications in Proposition 1.

Proposition 2 *1.a Holding VSL constant, an increase in π_ρ , the informativeness of the signal ρ_i , increases the gap in screening rates across risk groups: $\frac{\partial E[\theta_i|d_i=H]-E[\theta_i|d_i=L]}{\partial \pi_\rho} > 0$.*

1.b Holding VSL constant, an increase in π_ρ , the informativeness of the signal ρ_i , increases the gap in subjective risk across risk groups :

$$\frac{\partial E[E[\mu_i|\rho]|d_i=H]-E[E[\mu_i|\rho]|d_i=L]}{\partial \pi_\rho} > 0$$

2.a Holding π_ρ constant, an increase in VSL_i does not affect the gap in subjective risk across risk groups: $\frac{\partial E[E[\mu_i|\rho]|d_i=H]-E[E[\mu_i|\rho]|d_i=L]}{\partial VSL} = 0$.

2.b Depending on the distribution of c_i , an increase in VSL_i can increase or decrease the gap in screening rates across risk groups $\frac{\partial E[\theta_i|d_i=H]-E[\theta_i|d_i=L]}{\partial VSL} > 0$.

The proof of these propositions is straightforward and given in Appendix 1. This proposition is useful, because it both delivers testable implications on screening behavior and on subjective risk assessments that allow us to investigate whether the quality of information about cancer risks differs across a population. In the framework of this model, the hypothesis that educated individuals are better informed amounts to assuming that π_ρ increases in education. Proposition 1 tells us how this hypothesis can be tested.

First, according to 1.a, we expect those with better information to respond more in their screening behavior as objective risks vary. This implication is similar in spirit to the behavior investigated in a number of other papers: do those that we believe to hold better information react more strongly to differences in the productivity of treatments or the risk associated with behaviors? This question has for instance been investigated by de Walque (2004a) with respect to smoking behavior and the dissemination of new information on smoking risks. In a second paper, de Walque (2004b) tests this same implication by examining differences by education in the response to HIV epidemic in Uganda during the 1980s. We can also interpret Lichtenberg and Lleras-Muney (2002) as a test of this implication: the response of

the educated to new treatment options is generated by their advantage in adopting new technologies.

Unfortunately, proposition 2.b illustrates the limits of this approach. Even without an information advantage, the educated might respond more to differences in risk, simply because they are willing to expend more resources on health. This matters, because differences in the VSL across education groups (whether caused by education or simply due to spurious correlation) are the main competing explanation for the observed education-health gradient. It is therefore not possible to use data on health behavior such as screening alone to distinguish between the two main competing hypotheses: the efficiency hypothesis and the hypothesis of demand differences across education groups. These difference in screening behavior can only be used to distinguish the efficiency hypothesis from the demand hypothesis if we can fully control for all demand differences across education groups.⁸

The data from subjective risk assessments is therefore of particular interest. Propositions 1.b and 2.b show that we can use this data together with the objective risk data to test for differences in the quality of information across education levels. The data from subjective risk assessments has the advantage that differences in the quality of information by education is predicted to generate an interaction between education and objective risks in the subjective risk assessments (Proposition 1.b), while there is no reason to expect such an interaction because of differences in the VSL (Proposition 2.a).

4 Data

We are able to go beyond the existing empirical literature on the allocative efficiency hypothesis because of the rich data on subjective cancer risks,

⁸The same problem affects the papers by de Walque (2004a, 2004b) and Lichtenberg and Lleras-Muney (2002) discussed above as well as other papers in this literature. It is for this reason that Lichtenberg and Lleras-Muney (2002) control for income and other demand factors in their paper. Their results therefore depend on their ability to fully control for all possible differences in the VSL across education. In our empirical work, we will also include controls for determinants of the willingness-to-pay for health, but we are sceptical that we can fully adjust for differences in VSL across education groups.

objective cancer risks and cancer screening made available through the National Health and Interview Survey (NHIS). The NHIS is an annual household survey of the civilian, non-institutionalized population of the US. The NHIS records demographic and socioeconomic data as well as data on health behaviors, health status, and access to health care. In selected years additional modules are administered as part of the NHIS. The 2000 and 2005 Surveys include a cancer control module.

Our sample consists of non-Hispanic respondents to the cancer control module. We consider 3 cancers: breast cancer, cervical cancer, and colorectal cancer. These are cancers for which standard accepted screening methods were available in both 2000 and 2005 and for which we have screening data. Women aged 30-85 answered questions related to breast and cervical cancer and both males and females aged 40-85 answered questions related to colorectal cancers. After excluding individuals with missing data on crucial questions as well as those individuals that have already been afflicted by cancer⁹, we retain for each survey year between 6,487 and 10,944 observations depending on the gender and age-range.

The screening questions refer to mammograms for breast cancer, pap-smears for cervical cancer, and colonoscopies, sigmoidoscopies, and proctoscopies jointly for colorectal cancer. Individuals reported both whether they ever underwent a particular screen and how often they did so in the last 6 (10 for colorectal cancer) years. As shown in panel A of table 1, a large majority of the sample has been screened for breast cancer and even more so for cervical cancer. For cervical cancer specifically, this will limit the amount of information available from the question "ever screened?". However, there is a lot of variation in the intensity of screening and we will therefore model the intensity of screening using Tobit specifications to allow for censoring.

Statistics related to how objective risk factors are shown in panel B of table 1. We call these risk factors "objective", because they are based on scientific evidence and the advice of leading public health institutions. For breast cancer, there exists a well established medical model of risk, the Gail model. We can implement the Gail model using data on age, age at

⁹Details in the Data Appendix

menarche, age when first child was born, parity, race, and the number of direct relatives that was ever diagnosed with breast cancer.¹⁰ The Gail index represents a relative risk. We observe that the relative risk present in the population has increased somewhat over the 2000 to 2005. The later age of first birth, the earlier age of first menstruation and the increased presence of cancer in the family are all contributing to this trend. About 10% of women have at least one direct relative who has been diagnosed with breast cancer, but less than 1% of women have more than one relative with breast cancer.

The main risk factor for colorectal cancer in our data is the number of direct relatives who have been diagnosed with colorectal cancer. About 8% of the sample had 1 member and less than 1% had more than one member with colorectal cancer in the family.

We also have data on how much cervical cancer is present in the family, but it is important to note that cervical cancer in the family is not a risk factor for cervical cancer. The Center of Disease Control (CDC) for instance lists on its webpage a family history of cancer as a risk factor for both breast cancer¹¹ and colorectal cancer¹² but not for cervical cancer.¹³ This allows us to use the data on cervical cancer screening and its relation to the presence of cervical cancer in the family to conduct placebo tests. Cervical cancer is less common than breast or colorectal cancer and fewer women have relatives diagnosed with cervical cancer.

The subjective risk assessment questions differ for both survey years. For 2000, the question refers to the overall risk of developing cancer of any kind. About 55-60% of the population state that they believe themselves to be of low risk of developing cancer, whereas just more than 10% state themselves to be of high risk. These numbers are fairly stable across genders and age. The 2005 data contains measures of subjective risk for two specific cancers:

¹⁰The Gail model also uses information based on past screening results. Since the Gail model treats all components as independent, we can construct an average relative risk conditional on characteristics without using the past screening results. More detail on the Gail model is available in the Data Appendix.

¹¹http://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm

¹²http://www.cdc.gov/cancer/colorectal/basic_info/risk_factors.htm

¹³http://www.cdc.gov/cancer/cervical/basic_info/risk_factors.htm

breast cancer and colorectal cancer. The fraction of individuals believing themselves to be more likely to develop cancer than the average person lies around 10% and is lower for colorectal cancer than it is for breast cancer.

The remainder of table 1 displays socioeconomic variables used in the study. Noteworthy is that only a small share ($\sim 10\%$) of the sample does not have any health care coverage. Even among those less than 65 years old and thus not covered by medicare, the share without health care coverage is only about 13%. In 2005, the data details the type of insurance, which we use in our specifications.¹⁴ A final variable of note is the indicator whether a doctor recommended screening for breast cancer during the last 12 months available in 2005. For a large fraction of women, this was indeed the case, and we will use this variable in the specifications of breast cancer screening and subjective risk discussed next.

5 Empirical Results

We now turn to examine the predictions of our main proposition. We begin by examining whether education and objective risk factors interact positively to determine cancer screening and then turn to examine how education and objective risk factors interact in subjective risk assessments.

Table 2 shows estimated interactions between education and various risk factors for breast, colorectal, and cervical cancer in screening decisions. Each cell corresponds to an estimate from a different regression model.¹⁵ Reported are the point estimates and the standard errors for the Probits and Tobits. For the Probit estimates, we also present the average marginal effect as well as the average z-statistics for these interaction. We follow Ai and Norton (2003, 2004) in calculating the marginal effects and z-statistics accounting for the non-linearity of the Probit model. We use Probits to model whether an individual has ever undergone screening and use Tobits to model the frequency of screening in the last 6 years (10 for colorectal examinations).

¹⁴Whether individuals were enrolled in Medicare, Medicaid, single service health plans, private health insurance, state or federally sponsored health plans, military health insurance or indian health service.

¹⁵Complete regression results are available upon request.

We show estimates for each year separately and also for both years jointly. As controls we include a full set of age, income, gender (for colorectal) and race dummies. For the main effects we include dummies of education and the risk factor itself.

Consider first the cancers for which we have data on objective risk: breast and colorectal cancer. For both of these cancers, we find positive and interactions between the objective risk factor and education in all specifications and years. And, for all of these specifications with the exception of the 2005 Probits, we reject the one-sided null hypothesis of a negative interaction between risk and education at at least the 10% level. Indeed, in most cases we reject at the 1% level.

We do not expect and indeed do not find (with the exception of the 2005 Probit) the same statistically positive interaction if we consider cervical cancer screening. We do not expect this positive interaction because cervical cancer among immediate relatives is not considered a risk factor for cervical cancer. This suggests that our results for breast cancer and colorectal cancer are not driven by some other variable that correlates both with the incidence of cancers and with screening behavior.

Thus, we fail to reject the allocative efficiency hypothesis using screening behavior alone. However, this result is limited by the observation that differences in the VSL by education might also interact with objective risk to generate the same observed interaction. Indeed, we find it plausible that even if there is no information difference across education, the same difference in objective risk of developing cancer will generate a larger response among agents that have a higher willingness to pay for mortality reductions. Such a demand-based hypothesis would however not explain similar observed interactions of education and objective risk in subjective risk assessments of individuals, unless we allow for an information advantage for educated individuals. We therefore next examine how objective risk and education interact in subjective risk assessments.

Following the same format of table 2, we show in table 3 the estimated interactions of education with objective risk factors for the overall subjective risk of developing cancer from the NHIS 2000 as well as the cancer site-

specific subjective risks. The responses are recorded on a 3-point scale and we model these using an ordered Probit. We also report for each subjective risk the interactions with the number of relatives with cervical cancer. We predict that the coefficients for the Gail Index and the number of relatives with breast cancer are positive for the overall subjective cancer risk as well as for the breast cancer risk in 2005, but not for the colorectal cancer in 2005. Similarly, we predict that the interaction for the number of relatives with colorectal cancer is positive for the subjective overall cancer and colorectal cancer risks, but not for the breast cancer risk. Finally, we do not predict any particular sign for the interaction of the number of individuals with cervical cancer in either the overall or cancer site-specific subjective risks. All of these predictions are borne out in the data.

The qualitative patterns in the estimated coefficients reported in table 2 and 3 have a straightforward interpretation: the educated are better informed about the cancer risks they are facing. Because the educated are better informed about the risks they are facing, these risks are reflected both in their subjective risk assessments and in their screening behavior. This explanation has the main advantage of delivering a unified account of the patterns both in screening and in risk responses. We thus believe the evidence provided so far to provide strong support that the educated are in fact better informed about health risks and that this leads them to react to the risks they face by taking preventive action in the form of screening. In short, we believe the evidence favors the allocative efficiency hypothesis.

To find support for the allocative efficiency hypothesis is not same as rejecting the hypothesis that education differences in the demand for health determine health decisions. Evidence consistent with the demand hypothesis is that both income and education enter very strongly into the screening decision in all our specifications, including those that control for subjective risk.¹⁶ In table 4 and 5, we ask whether there is evidence that such education demand differences for health generate the observed interactions between risk and education in screening decisions. First, we ask whether

¹⁶The full results for all our specifications are available upon request – we are not showing all of these estimates here in order to save space.

controlling for subjective risks eliminates the interaction between education and risk factors in screening decisions. If it does not, then this would make it plausible that at least some of the interaction between risk and protective action is due to other demand factors correlated with education rather than just information. In table 4, we use the pooled 2000 and 2005 data to show that controlling for subjective risk does attenuate, but not eliminate the interactions. Table 5 reports similar results using the 2005 data only, since for 2005 we have the site-specific risk assessments. We find¹⁷ again that the subjective risk attenuates, but does not eliminate the observed education-risk interaction in screening behavior for mammograms. However, for colorectal screening the role of information is large and once we control for subjective risks, there risk-education interaction disappears.

It is plausible that family income is an important driver of the overall demand for health. We therefore investigate the demand hypothesis by examining whether the estimated risk-education interactions are robust to including interactions of income with the risk factors. If income is a good proxy for the demand for health and if the demand for health drives the observed screening interactions, then interacting income with education should eliminate the interaction between risk and education. Using the mid-points of the 13 categories in which income is reported as a multiple of the federal poverty line, we construct a continuous income measure that we interact with the various risk factors. In addition, we interact the dummy that indicates that incomes exceed 5 times the poverty line with the risk indicators. The evidence in table 5, columns 3, 7, and 11 shows that the education-risk interaction remains positive and statistically significant for both breast cancer and colorectal cancer screening. Thus, we find little evidence that the observed risk and education interactions are due to a spurious correlation of education with the overall demand for health.

We conclude the empirical analysis of the NHIS data by examining the role of doctor recommendations in the decision to screen. This is possible for the case of breast cancer screening because the NHIS 2005 records whether women were recommended to undergo a mammogram by a doctor

¹⁷Columns 1 and 2, 5 and 6, and 9 and 10 of table 5

within the last year. In table 6, we examine whether the positive estimated risk-education interaction is robust to the inclusion of this variable both in the screening and the subjective risk equations. The recommendation of a doctor to undergo a screen is both an important predictor of the screening decision itself and it also predicts whether a women perceives herself to be of higher cancer risk. However, it does not eliminate the positive relationship between the risk factor and education in either the subjective risk assessments nor the screening decision. This is largely due to the fact that there the interaction between risk and education in the estimated model of doctor recommendations themselves is small: doctors are not significantly more likely to tailor their screening recommendations to the actual patient risk among the educated than the less educated.

Overall, the reported coefficient patterns are therefore consistent with the implications of the allocative efficiency model stated in 1.a. and 1.b. from proposition 1. We find these consistent patterns with respect to the risk factor "relatives with cancer" only for those cancers where this is in fact an accepted risk factor. Furthermore, we find these patterns to be robust to allowing for an independent demand effect through the inclusion of income interacted with risk in the specifications and we find little evidence that the interaction simply reflects doctor recommendations. Overall, the patterns are consistent with the notion that educated individuals might inform themselves about what it means for their own cancer risk if an immediate family member develops breast, cervical, or colorectal cancer. And, they are more likely to indeed act on this information.

6 Educated to Believe in Science: A Potential Mechanism

In the NHIS data, educated individuals who face a higher risk of developing cancer are disproportionately more likely to both state that they perceive themselves to be of higher risk and to undergo cancer screening. This is true even after controlling for income and access to health care providers and is

robust to a variety of different specifications. Furthermore, controlling for subjective risk assessments attenuates the relation between screening behavior and objective risks. The allocative efficiency hypothesis provides a simple explanation for these facts: the educated are better informed about the presence of risk factors than the less educated and they use this information to make better informed screening decisions.

The question then arises how education might improve information? It is unlikely that agents literally receive signals on their cancer risk and that the frequency of this signal differs with education. These assumptions are useful for analytic purposes, but they hardly describe reality. How then can education positively affect health care decision making? We propose that the educated might be generally more positively inclined towards a scientific world-view. By contrast, the less educated might be more likely to hold non-scientific beliefs about what determines their life-experiences, leading them to discount the findings of the medical science and thus limiting their ability to benefit from modern medicine.

We are not the first to suggest that differential beliefs and faiths in the medical system may explain health disparities. For instance, in their contribution on the adoption of medical technologies, Lichtenberg and Lleras-Muney (2002) cite survey evidence from the National Science Foundation that "53% of those with more than a college degree understand the nature of scientific inquiry, whereas only 4% of those with less than a high school degree do"¹⁸ and that those with college degrees are more interested in medical discoveries and more inclined to believe that the benefits of new technologies outweigh the costs.

We present new evidence from cancer screening that further support the hypothesis that education leads to more favorable attitudes towards the medical sciences. The data presented here comes from a 2002 survey conducted by the San Francisco Mammography Registry of 1,700 women screened for breast cancer. The survey was designed to examine the views of women towards various aspects of breast health, including their attitudes

¹⁸Science and Engineering Indicators 2000, published by the National Science Foundation: <http://www.nsf.gov/sbe/srs/seind00/access/toc.htm>

towards screening. Clearly this survey is not representative of the general US population. The sample is choice based (only women who choose to screen were selected into the sample) and it is drawn from a sample of women in San Francisco, hardly a representative community in the US.

Nevertheless, we believe that the results from this data are highly suggestive. Among other questions, the survey respondents were asked whether they agree with the following statement: "If a person prays about it, God will protect her from getting cancer." The fraction of respondents increases from about 8% among those with more than a high school degree to 22% among women who completed high school but did not continue with education to 50% among respondents with less than a high school degree. This relation persists if we adjust for age and income as well as marital status. Within this sample the less educated were also more likely to hold non-scientific beliefs with respect to breast cancer risk. More generally, they were more likely to display a fatalistic attitude towards health. For instance, only 27% of those with more than a high school degree agree that "If it is meant to be, I will stay healthy". 49% of those with only a high school degree agree with this statement and 69% of those with less than a high school degree displayed this fatalistic attitude.

The survey also contains data on the breast cancer risk faced by women. We constructed an index of whether a women perceived herself to be of greater or lower than average breast cancer risk. This index was based on 4 questions that asked women to rate their breast cancer risk during their life-time and compare it with that of average women their age. We then related this index of self-rated risk to the Gail index using a simple linear regression. The errors from this regression allowed us to crudely classify women into those who over-estimate their risk and those who underestimate their risk of developing breast cancer.

We use this classification to test whether women who underestimate their breast cancer risk are disproportionately likely to also agree with the statement that prayer protects from cancer. This was roughly confirmed in the data. According to the simple index above, 56% of women who believe in the protective power of prayer also underestimate their cancer

risk, whereas only 44% of those who do not believe in the protective power of prayer underestimated their risk. Together these findings suggest that less educated women are indeed less likely to hold scientific beliefs about the causes of breast cancer and that these non-scientific beliefs result in underestimating the personal risk from breast cancer.

7 Conclusion

In this paper, we have presented new evidence from cancer screening and from subjective cancer risks that supports the allocative efficiency hypothesis. Our results are novel in that they link individual risk assessments and individual health care decisions to objective risk factors. We develop two tests of the allocative efficiency hypothesis that can be implemented using publicly available data. We find that both screening and individual risk assessments are more closely related to objective risk factors among educated individuals. The simplest explanation for these patterns is that the educated are better at processing health information and that this confers an advantage in making health decisions.

Our study does have two important limitations. First, the data analyzed in this paper does not allow us to determine whether education causes individuals to become better at processing information or whether education simply correlates with the ability to process information. We cannot rule out an ability bias that might lead us to observe better health choices among the educated even though education might not itself improve health decision making in a causal sense.

Second, it is possible that the observed information advantage with education is the outcome of a prior investment decision. In that case, it might be that the educated know more about their cancer risk, but only because they expended more resources to learn about health. In that case, the information advantage observed when screening decisions are made reflects simply a difference in the demand for health that has led to more investments into information at an earlier state.

We do not believe that the current data allows us to address these two

questions. For now, our evidence simply suggests that the educated indeed possess an information advantage and thus supports a crucial assumption of the allocative efficiency hypothesis. Better data is required to make further progress.

8 References

Auster, R., I. Leveson and D. Saracheck (1969), "The production of health: an exploratory study", *Journal of Human Resources* 4: 411-436.

Cutler, D. and A. Lleras-Muney (2007), "Understanding Differences in Health Behaviors by Education." Manuscript, Dept. Econ., Univ. California, Los Angeles

Cutler, D. and A. Lleras-Muney (2008), "Education and Health: Evaluating Theories and Evidence", In *Making Americans Healthier: Social and Economic Policy as Health Policy*, edited by J. House, R. Schoeni, G. Kaplan, and H. Pollack, The Effects of Social and Economic Policy on Health, New York: Sage.

Cutler, D. Lange, F, Meara, E., Richards, S., and C. Ruhm (2010) "Explaining the Rise in Educational Gradients in Mortality", manuscript available at <http://www.econ.yale.edu/~fl88/DisparitiesTrends.pdf>

De Walque, D. (2004a), "Education, information, and smoking decisions: evidence from smoking histories, 1940-2000" Policy Research Working Paper Series 3362, The World Bank.

De Walque, D. (2004b), "How Does the Impact of an HIV/AIDS Information Campaign Vary with Educational Attainment? Evidence from Rural Uganda" Policy Research Working Paper Series 3289, The World Bank.

Erbsland, M., W. Ried and V. Ulrich (1995), "Health, health care, and the environment. Econometric evidence from German micro data", *Health Economics* 4: 169-182.

Fletcher, J.M., and D.E. Frisvold (2009), "Higher Education and Health Investments: Does more Schooling Affect Preventive Health Care Use?", *Journal of Human Capital*, vol. 3, no.2.

Fuchs, V. R. (1982), "Time preferences and health: an exploratory

study", in : V. R. Fuchs, ed. *Economic Aspects of Health* (University of Chicago Press, Chicago) 93-120.

Gail M. H., et al. (1989) "Projecting individualized probabilities of developing breast cancer for white females who are being examined annually" *J Natl Cancer Inst* 81: 1879-86.

Gail, M.H., et al. (1999) "Weighing the Risks and Benefits of Tamoxifen Treatment for Preventing Breast Cancer" *Journal of the National Cancer Institute*, Vol. 91, No.21.

Gilleskie, D.B. and A.L. Harrison (1998), "The effect of endogenous health inputs on the relationship between health and education", *Economics of Education Review* 17: 279-297.

Goldman, D.P. and D. Lakdawalla (2001), "Understanding Health Disparities Across Education Groups", NBER Working Paper 8328.

Goldman, D.P. and J.P. Smith (2002), "Can patient self-management help explain the SES health gradient?" *Proceedings of the National Academy of Sciences of the United States* 99:10929-10934.

Grossman, M. (1972a), "On the concept of health capital and the demand for health", *Journal of Political Economy* 80: 223-255.

Grossman, M. (1972b), *The Demand for Health: A Theoretical and Empirical Investigation* (Columbia University Press for the National Bureau of Economic Research, New York).

Grossman, M. (2000), "The human capital model", in: A.J. Culyer and J.P. Newhouse, eds, *Handbook of Health Economics* (Elsevier, Amsterdam) Chapter 7.

Grossman, M., and R. Kaestner (1997), "Effects of education on health", in J.R. Behrman and N. Stacey, eds., *The Social Benefits of Education* (University of Michigan press, Ann Arbor, MI) 69-123.

Kenkel, D..S. (1991a), "Health behavior, health knowledge, and schooling", *Journal of Political Economy* 99(2): 287-305.

Kenkel, D..S. (1991b), "What you don't know really won't hurt you", *Journal of Policy Analysis and Management* 10(2): 304-309.

Kenkel, D. S. (1994), "The demand for preventive medical care", *Applied Economics* 26: 313-325.

Kenkel, D. S. (2000) "Prevention.", In *Handbook of Health Economics*, edited by A. Culyer and J. Newhouse. Amsterdam: Elsevier.

Leigh, J.P. (1990), "Schooling and seatbelt use", *Southern Economic Journal* 57: 195-210.

Lichtenberg, F and A. Lleras Muney (2002), "The Effect of Education on Medical Technology Adoption: Are the More Educated More Likely to Use New Drugs?" NBER Working Paper 9185.

Manski, Richard J. and Magder, L.S. (1998). "Demographic and Socioeconomic Predictors of Dental Care Utilization," *Journal of the American Dietetic Association*, 129, 195.

Meara, E., Richards, S. and D. Cutler (2008) "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, By Education, 1981-2000.) *Health Affairs (Millwood)* 27:350-360.

Mullahy, J. (1999), "It'll only hurt a second? Microeconomic determinants of who gets flu shots", *Health Economics* 8(1): 9-24.

Michael, R. T. (1972), *The Effect of Education on Efficiency in Consumption* (Columbia University Press for the National Bureau of Economic Research, New York).

Michael, R.T, and G.S. Becker (1973), "On the new theory of consumer behavior", *Swedish Economic Journal* 75: 378-396.

Murphy, K. and R. Topel (2005), "The Value of Health and Longevity" NBER Working Paper. No. W 11405.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition" *The Journal of Political Economy*, Vol. 82, No. 1, pp 34-55.

Schoenbaum, M. (1997), "Do smokers understand the mortality effects of smoking? Evidence from the health and retirement survey", *American Journal of Public Health* 87(5): 755-759.

Viscusi, K.W. (1990), "Do smokers underestimate risks?", *Journal of Political Economy* 98(5): 1253-1269.

Viscusi, K.W. (1993), "The Value of Risks to Life and Health", *Journal of Economic Literature*, Vol. XXXI (Dec 1993), pp. 1912-1946.

Viscusi, K.W. (1998), "Public perception of smoking risks", Paper pre-

pared for the International Conference on the Social Costs of Tobacco, Lausanne, Switzerland, August 21-22, 1998.

Wagstaff, A. (1986), "The demand for health: some new empirical evidence", *Journal of Health Economics* 5: 195-233.

Wagstaff, A. (1993), "The demand for health: an empirical reformulation of the Grossman model", *Health Economics* 2: 189-198.

9 Appendix 1: Proof of Proposition 1

To proof the four part of proposition 1, we begin with implication 2.b. and thus need to show how subjective risk of cancer differs with the objective risk d_i . From Bayes rule and some algebraic manipulation we get:

$$E[E[\mu|\rho_i]|d_i = 0] = \mu_0 + (\mu_1 - \mu_0) * \frac{(1 - \pi_\rho) P(d = 1)}{1 - \pi_\rho P(d = 1)} \quad (\text{A.1})$$

and because $E[E[\mu|\rho]] = E[\mu]$:

$$E[E[\mu|\rho_i]|d_i = 1] = \frac{E[\mu] - E[E[\mu|\rho_i]|d_i = 0] \Pr(d = 0)}{\Pr(d = 1)} \quad (\text{A.2})$$

All components in equation (A.1) are primitives to the model and we can therefore differentiate with respect to π_ρ . This reveals directly that an increase in the quality of the signal (π_ρ) increases the difference in subjective risk across risk classes d_i : $\frac{\partial E[E[\mu|\rho_i]|d_i=1] - E[E[\mu|\rho_i]|d_i=0]}{\partial \pi_\rho} > 0$ and therefore proves 1.b.

Consider now implication 1.a. Conditional on the risk class (and implicitly conditioning on VSL_i), the probability of screening is

$$\begin{aligned} \Pr(\theta_i = 1|d_i) &= \Pr(\rho_i = 1|d_i) * \Pr(\theta_i = 1|\rho_i = 1) + (1 - \Pr(\rho_i = 1|d_i)) * \Pr(\theta_i = 1|\rho_i = 0) \\ &= \Pr(\theta_i = 1|\rho_i = 0) + \pi_\rho^{1(d_i=H)} (\Pr(\theta_i = 1|\rho_i = 1) - \Pr(\theta_i = 1|\rho_i = 0)) \end{aligned} \quad (\text{A.3})$$

and

$$\Pr(\theta_i = 1|d_i = H) - \Pr(\theta_i = 1|d_i = L) = \pi_\rho (\Pr(\theta_i = 1|\rho_i = 1) - \Pr(\theta_i = 1|\rho_i = 0)) \quad (\text{A.4})$$

Now, using eq (2) and $\Pr(c_i = 1|\rho_i) = E[\mu|\rho_i]$

$$\theta_i = 1 \Leftrightarrow c_i \leq E[\mu|\rho_i] gVSL_i \quad (\text{A.5})$$

Because $\frac{\partial \Pr(d_i=1|\rho_i=0)}{\partial \pi_\rho} < 0$ and $E[\mu|\rho_i] = \mu_0 + (\mu_1 - \mu_0) \Pr(d_i = 1|\rho_i)$ we get $\frac{\partial E[\mu|\rho_i=0]}{\partial \pi_\rho} < 0$. Furthermore, $E[\mu|\rho_i = 1] = \mu_H$. It is immediate from eq. (A.4) that $\Pr(\theta_i|\rho_i = 0)$ declines in π_ρ and that $\Pr(\theta_i|\rho_i = 1)$ is constant with respect to π_ρ . Therefore the difference in screening behavior across risk classes as shown in equation (A.4) increases with π_ρ as stated in Implication 1.a.

The claim in implication 2.b that the willingness-to-pay does not affect the subjective risk assessment requires no proof. However, we still need to show that the increases in the willingness-to-pay with schooling can raise the probability of screening. From equation (A.5) we have that:

$$\begin{aligned} & \frac{\partial \Pr(\theta_i = 1|d_i = H) - \Pr(\theta_i = 1|d_i = L)}{\partial VSL_i} \quad (\text{A.6.}) \\ = & \pi_\rho (f(E[\mu|\rho_i = 1] gVSL_i) * gE[\mu|\rho_i = 1] - f(E[\mu|\rho_i = 0] gVSL_i) * gE[\mu|\rho_i = 0]) \end{aligned}$$

where $f(\cdot)$ is the density of c_i . This expression can not be signed without imposing additional restrictions on $f(\cdot)$. This completes the proof.

10 Appendix 2: Data Appendix

In its 2000 and 2005 cancer screening modules, the NHIS surveyed individuals on their cancer screening behavior and asked them questions about what risk to develop cancer they think they face. In addition, the NHIS contains a number of objective risk factors, that is variables that are medically shown to predict the incidence of cancers. Most important among these objective

risk factors is the number of direct relatives afflicted with particular types of cancers.

We consider 3 cancers: breast cancer, cervical cancer, and colorectal cancer. These are cancer for which standard accepted screening methods were available in both 2000 and 2005 and for which we have screening data. Female respondents aged 30-85 have been asked about whether they ever had a mammogram performed and how many they had performed within the last 6 years. They were asked similar questions about pap-smears, the main test for cervical cancer. Both males and females aged 40-85 were asked if they ever receive a colonoscopy, sigmoidoscopy, or proctoscopy and how many of these they underwent in the last 10 years.

We use the Gail index, a standard medically accepted measure of breast cancer risk to measure the objective risk factors faced by individuals. The Gail Index is a function of the family history of cancer, of the number of positive screening results a woman has had so far, of the age at menarche, of the fertility history of a woman, and of her age. Table A.1 summarizes how one can construct the Gail index to construct individual relative cancer risk. The information about breast cancer risk summarized by the Gail model goes beyond a list of risk factors and also includes the size of the impact of these risk factors and the interaction with other risk factors.

Table A.1

Table A.1 shows how to obtain the relative risk of developing breast cancer within the next 5 years relative to a baseline risk of developing breast cancer that varies with age. The relative risk is obtained by multiplying four contributions to relative risk from categories A,B,C and D. In order to obtain the actual 5-year risk of developing breast cancer, one multiplies this relative risk with an age specific baseline risk that increases from 0.134% at age 30-34 to 1.157% at age 70-74 and then declines slightly to 1.006% at age 80-84/

The past history of cancer screening results is clearly endogenous. For that reason we use an abridged version of the Gail index that measures the

cancer risk of a woman conditional on those risk factors that do not depend on screening itself. This is the appropriate measure of risk if the screening decision is measured using the answer to "have you ever been screened for breast cancer". The variables required to construct the abridged Gail index are parity, age at first birth, age at menarche and the number of direct relative that have developed breast cancer. We use these variables to generate the Gail Index. In order to test for sensitivity to not using the history of screening results, we repeat our analysis for women both across the entire age range and for women aged 30-60 only. Since this latter group has received substantially fewer screens, the likelihood of having a positive screen is reduced and the abridged Gail Index provides a better proxy for breast cancer risk. All our results are consistently found in the sample consisting of younger women only.¹⁹

10.0.1 The National Health and Interview Survey

The NHIS is an annual household survey of the civilian, non-institutionalized population of the US. The NHIS records demographic and socioeconomic data as well as data on health behaviors, health status, and access to health care. In selected years additional modules are administered as part of the NHIS. The 2000 and 2005 Surveys include a cancer control module.

In both survey years about 40,000 families with a total of 100,000 family members were interviewed. In each household one adult (the "sample adult") and one child (the "sample child") are asked a more detailed set of questions. In 2000 (2005) there were 32,374 (31,428) Sample Adults. Consider the selection of the female sample of women aged 30-85. The other samples are comparable. We are limiting ourselves to non-Hispanic sample adult females aged 30-85. This leaves us with 11,764 (11,726) women aged 30-85 in 2000 (2005). Dropping individuals with invalid answers about education, whether they ever had cancer, and on whether they have had a mammogram removes 75, 6, and 271 (2005: 125/13/871) observations respectively. A further 335 (369) women report having had breast cancer and

¹⁹A more complete analysis that incorporates the past screening history requires a dynamic model of screening that is beyond the current scope of the analysis.

are likewise dropped. In order to construct the Gail Index we require the age of onset of menstruation, information on whether a woman has ever given a live birth and at what age, and also the number of direct family members (parents, siblings, and children) who have ever developed breast cancer. Insufficient or incoherent responses for these variables removes another 698 (680) individuals. We thus retain 10,379 (9,668) women in the appropriate age range. The data appendix describes the sample selection in more detail.

For the screening decision, our dependent variables are an indicator variable describing whether the woman had ever undergone a mammogram and as an alternative independent variable the number of mammograms the woman has received during the last 6 years. This variable allows us to examine how the intensity of cancer screening varies across individuals. For subjective risk assessments, we have categorical variables describing the subjective risk of individuals. The content of these variables differs between 2000 and 2005. In 2000, we have a categorical variable (low, medium, high) describing the subjective overall cancer risk, and in 2005 we have a similar variable describing the subjective breast cancer risk.

We use as additional controls such as education, a categorical variable on family income (relative to the poverty line), the size of the MSA the woman resides in, and various variables describing health care coverage (medicare, private, etc...). Finally, we are also using a variable from the 2005 NHIS that indicates whether a woman has been counselled by her physician to receive a mammogram. We use this variable both as a dependent variable and as a control.²⁰

²⁰In 2000 this variable was unfortunately only administered to those women who have not been screened previously.

Table 1 Summary Statistics for 2000 and 2005 National Health and Interview Survey

	2000			2005		
	Female 30-85	Female 40-85	Male 40-85	Female 30-85	Female 40-85	Male 40-85
Observations	10,944	8,288	6,487	10,198	8,066	7,038
Panel A: Screening Outcomes						
% that ever had:						
a Mammogram	72.6	86.0	na	75.3	87.2	na
a Pap Smear	97.0	96.5	na	97.2	97.2	na
a Colorectal Exam	na	30.4	32.3	na	39.8	39.8
Conditional on any, frequency of :						
Mammograms within 6 years	3.92 (2.23)	4.14 (2.18)	na	4.01 (2.19)	4.21 (2.14)	na
Pap-smears within 6 years	4.84 (2.22)	4.67 (2.15)	na	4.61 (2.18)	4.46 (2.12)	na
Colorectal Exams within 10 years	na	1.86 (1.87)	2.14 (2.09)	na	1.93 (1.76)	1.93 (1.76)
Panel B: Risk Factors						
Gail Index ²⁾	1.03 (0.78)			1.09 (0.82)		
Parity > 0	81%			81%		
Age at First Birth (if parity>0)	23.01 (4.99)			23.24 (5.14)		
Age of first menstruation	12.83 (1.77)			12.76 (1.70)		
# of direct relatives with						
breast cancer: 0	90.15%			88.94%		
1	9.01%			10.10%		
>1	0.84%			0.96%		
col. cancer: 0		92.18%	93.43%		91.54%	93.89%
1		7.22%	6.01%		7.72%	5.63%
>1		0.6%	0.55%		0.37%	0.48%
cerv. cancer: 0	97.49%			97.16%		
1	2.30%			2.61%		
>1	0.21%			0.04%		

Cont'd on next page.

	2000	2000	2000	2005	2005	2005
	Female	Female	Male	Female	Female	Male
	30-85	40-85	40-85	30-85	40-85	40-85
Panel C: Subjective Risk Assessments						
Subjective Cancer Risk³⁾						
Overall Cancer Risk Low:	56.46%	59.00%	57.81%			
Medium:	31.42%	29.58%	31.65%			
High:	12.12%	11.42%	10.53%			
Breast Cancer Risk: “Less Likely”				36.93%	39.51%	
“About as Likely”				50.85%	48.90%	
“More Likely”				12.22%	11.59%	
Col. Cancer Risk: “Less Likely”					43.04%	39.34%
“About as Likely”					48.07%	52.87%
“More Likely”					8.90%	7.79%
Panel D: Socioeconomic Variables						
Education	13.22	13.01	13.31	13.50	13.34	13.57
	(2.48)	(2.54)	(2.73)	(2.46)	(2.40)	(2.62)
Hispanic	18.29%	16.90%	13.60%	17.64%	16.37%	14.51%
Age	53.23	59.19	57.47	53.98	59.13	58.00
	(15.89)	(13.60)	(12.62)	(15.48)	(13.21)	(12.59)
No Health Coverage	9.75%	7.80	9.08%	8.82%	8.34%	9.05%
Family income as multiple of poverty line ³⁾						
if < 5 * poverty line ⁴⁾	2.39	2.34	2.60	2.39	2.39	2.58
	(1.29)	(1.28)	(1.26)	(1.27)	(1.26)	(1.27)
% >5 * poverty line ⁴⁾	19.98%	19.87%	26.44%	21.12%	20.62%	26.97%
Doc Recommended Screening for Breast Cancer ⁵⁾	na	na	na	52.74%	59.64%	na

1) Reported are averages and standard deviations for the main variables and three samples analyzed: 30-85 year old women for breast cancer and cervical cancer screening and 40-85 year old men and women for colorectal cancer screen.

2) The construction of the Gail Index is described in the data appendix.

3) In 2000, the question refers to cancer risk in general and asks about “low, medium, high” risk. In 2005 a separate question is asked about breast and colorectal cancer and asks about relative risk.

4) Family income is a categorical variable with 13 categories for family income as a multiple of the federal poverty limit. This variable is top-coded at 5 times the federal poverty line.

5) Only available in 2005 cancer control module.

Table 2 Screening and Objective Risk¹⁾

	2000		2005		Both Years	
	(1)	(2)	(3)	(4)	(5)	(6)
Education Interacted with:	Probit	Tobit	Probit	Tobit	Probit	Tobit
Gail Index	0.032** (0.009) 0.007 (2.667)	0.081** (0.016) . (.)	0.044** (0.010) 0.008 (2.645)	0.083** (0.016) . (.)	0.037** (0.007) 0.007 (3.795)	0.081** (0.011) . (.)
# Relatives with Breast Cancer	0.066** (0.020) 0.013 (2.441)	0.081** (0.032) . (.)	0.012 (0.023) 0.000 (-0.304)	0.095** (0.032) . (.)	0.045** (0.014) 0.008 (1.868)	0.090** (0.023) . (.)
# Relatives with Colorectal Cancer	0.032** (0.013) 0.013 (3.236)	0.057* (0.035) . (.)	0.019* (0.015) 0.007 (1.691)	0.045* (0.028) . (.)	0.026** (0.010) 0.011 (3.589)	0.050** (0.022) . (.)
# Relatives with Cervical Cancer ²⁾	0.024 (0.088) 0.000 (-0.133)	0.002 (0.069) . (.)	0.159* (0.089) 0.008 (1.114)	0.063 (0.060) . (.)	0.080 (0.058) 0.003 (0.674)	0.034 (0.045) . (.)

1) Each cell reports estimates of the interaction of education with the relevant risk factor from a different Probit or Tobit regression for the indicated years. Reported are 4 numbers for Probits and 3 for Tobits. The top two numbers are the point estimate and the standard error (in parentheses) of the point estimate. For the Probits I also report the average marginal effect and the average z-score (in parenthesis) for the marginal effects. The marginal effects and the z-scores are calculated accounting for non-linear models using the method described in Ai and Norton (2003) and Ai and Norton (2004). For tobit models, we simply report the point estimates and standard errors for the linear index. Two stars (**) indicates that we reject a negative coefficient at the 1 % level, one star (*) at the 1% level, and a cross(+) at the 10 percent level.

The estimates in the first two rows of results refer to breast cancer screening, those in the third row to colorectal examinations and those in the fourth row to pap smears (cervical cancer screening). The Probit specifications have as their dependent variable whether an individual has ever received a screen of the specified type and the estimates in the Tobit specifications refer to the number of screens received in the last 6 years for mammographies and pap smears and 10 years for colorectal examinations.

All specifications control for a full set of age, health care coverage, income bracket, gender (for colorectal screening) and education indicator and also contains a risk-factor main effect. The specifications for colorectal cancer are estimated using both males and females age 40-85. The breast cancer and the cervical cancer models are estimated on females aged 30-85.

2) Contrary to breast or colorectal cancer, the number of direct relatives with cervical cancer is not a risk factor for cervical cancer.

Table 3 Subjective Risk Assessments and Objective Risk

	<i>Subjective Cancer Risk</i>	<i>Subjective Breast Cancer Risk</i>	<i>Subjective Colorectal Cancer Risk</i>
Education Interacted with:	2000	2005	2005
Gail Index	0.020** (0.006)	0.032** (0.006)	0.004 (0.006)
# Relatives with Breast Cancer	0.061** (0.013)	0.098** (0.014)	-0.006 (0.010)
# Relatives with Colorectal Cancer	0.011 (0.012)	0.021 (0.016)	0.080** (0.012)
# Relatives with Cervical Cancer	0.017 (0.029)	-0.006 (0.025)	-0.001 (0.021)

1) Each cell reports estimates of the interaction of education with the relevant risk factor from a different ordered Probit regression for the indicated years. In parentheses we report standard errors. Two stars (**) indicates that we reject the one-side test of the hypothesis that the estimate is negative at the 1%-level, 1 star (*) at the 5%-level, and a cross (+) significance at the 10%-level.

The estimates in the first two rows of results refer to breast cancer screening, those in the third row to colorectal examinations and those in the fourth row to pap smears (cervical cancer screening). The Probit specifications have as their dependent variable whether an individual has ever received a screen of the specified type and the estimates in the Tobit specifications refer to the number of screens received in the last 6 years for mammographies and pap smears and 10 years for colorectal examinations.

All specifications control for a full set of age, health care coverage, income bracket, gender (for colorectal screening) and education indicator and also contains a risk-factor main effect. The specifications for colorectal cancer are estimated using both males and females age 40-85. The breast cancer and the cervical cancer models are estimated on females aged 30-85.

2) Contrary to breast or colorectal cancer, the number of direct relatives with cervical cancer is not a risk factor for cervical cancer.

TABLE 4: SUBJECTIVE RISK AND SCREENING DECISIONS¹⁾

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	# of Mammograms in last 6 years				# of CREs in last 10 years		# of Pap Smears in last 6 years ²	
Educ*Risk	0.076** (0.011)	0.066** (0.011)	0.085** (0.023)	0.062** (0.023)	0.050** (0.022)	0.031+ (0.022)	0.001 (0.069)	0.000 (0.069)
2000 Subject. Risk Medium		0.321** (0.000)		0.289** (0.066)		0.350** (0.069)		0.214** (0.062)
2000 Subject. Risk High		0.693** (0.094)		0.600** (0.095)		1.149** (0.097)		0.287** (0.088)
2005 Subject. Risk Medium		0.124+ (0.067)		0.112+ (0.067)		-0.043 (0.063)		Na
2005 Subject. Risk High		1.025** (0.102)		0.885** (0.104)		1.584** (0.104)		Na
Risk Factor	Gail Index		# of relatives with breast cancer		# of relatives with colorectal cancer		# of relatives with cervical cancer	
Observations	20,774	20,774	20,774	20,774	28,864	28,864	10,637	10,637
Pseudo R-sq.	0.110	0.112	0.111	0.113	0.066	0.072	0.0417	0.0417

1) Reported are coefficients from a Tobit on the indicated variables for pooled 2000 and 2005 data. The estimates for cervical cancer screening (columns 7 and 8) are estimated on 2000 data only, because no specific risk assessment for cervical cancer is available in 2005. Standard errors are reported in parentheses. For the education and risk interaction ** indicates that we reject the one-side test of negative coefficient at the significance at the 1% , * at the 5% and + at the 10% level. For the other coefficients these symbols refer to the usual two-sided tests.

All specifications control for a full set of age, health care coverage, income bracket, gender (for colorectal screening) and education indicators and also contains a risk-factor main effect. The specifications for colorectal cancer are estimated using both males and females age 40-85. The breast cancer and the cervical cancer models are estimated on females aged 30-85.

The subjective risk question in 2000 refers to cancer risk from any cancer. In 2005, the subjective question refers to breast cancer and colorectal cancer specifically.

2) The 2005 data does not contain a subjective risk variable for cervical cancer. Columns 7 and 8 are therefore estimated on 2000 data only.

TABLE 5) CAN WE EXPLAIN THE EDUCATION-RISK INTERACTION WITH DEMAND DIFFERENCE FOR HEALTH WITH EDUCATION?¹⁾

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
	# of Mammograms in last 6 years						# of CREs in last 10 years						
Educ*Risk	0.080** (0.016)	0.067** (0.016)	0.060** (0.017)	0.049** (0.017)	0.094** (0.032)	0.063* (0.032)	0.098** (0.035)	0.071** (0.035)	0.044+ (0.028)	0.009 (0.027)	0.043+ (0.031)	0.014 (0.030)	
2005 Subject. Risk Medium		0.125+ (0.067)		0.125+ (0.067)		0.111+ (0.000)		0.111* (0.066)		-0.028** (0.055)		-0.026 (0.055)	
2005 Subject. Risk High		1.006** (0.103)		0.997** (0.103)		0.858** (0.107)		0.861** (0.000)		1.504** (0.093)		1.507** (0.093)	
Income*Risk			0.052** (0.016)	0.047** (0.015)			0.033 (0.034)	0.030** (0.034)			-0.049 (0.031)	-0.059 (0.031)	
Rich*Risk			0.543** (0.167)	0.483** (0.167)			0.045 (0.359)	-0.038** (0.358)			-0.172 (0.320)	-0.321 (0.315)	
Risk Factor		Gail Index			# of relatives with breast cancer			# of relatives with colorectal cancer					
Observations	10,034	10,034	10,034	10,034	10,034	10,034	10,034	10,034	10,034	14,414	14,414	14,414	14,414
Pseudo R-sq.	0.107	0.109	0.107	0.110	0.108	0.110	0.108	0.110	0.076	0.085	0.076	0.085	

1) Reported are coefficients from a Tobit on the indicated variables using the 2005 data. We focus on the 2005 data, because in 2005 we both have cancer-site specific subjective risks as well as data on doctor recommendations. Standard errors are reported in parentheses. For the education and risk interaction ** indicates that we reject the one-side test of negative coefficient at the significance at the 1% , * at the 5% and + at the 10% level. For the other coefficients these symbols refer to the usual two-sided tests. All specifications control for a full set of age, health care coverage, income bracket, gender (for colorectal screening) and education indicators and also contains a risk-factor main effect. The specifications for colorectal cancer are estimated using both males and females age 40-85. The breast cancer models are estimated on females aged 30-85.

TABLE 6: THE ROLE OF DOCTOR RECOMMENDATIONS¹⁾

	<i># of Mammograms in last 6 years</i>		<i>Subjective Breast Cancer Risk</i>		<i>Doctor Recommendation</i>	
Education *	0.071**	0.091**	0.034**	0.097**	0.013*	0.010
Risk	(0.015)	(0.030)	(0.007)	(0.014)	(0.007)	(0.015)
Doctor Rec	2.352**	2.343**	0.093**	0.079**		
	(0.059)	(0.059)	(0.026)	(0.027)		
Risk Factor	Gail Index	# Relatives with Br.Can	Gail Index	# Relatives with Br.Can	Gail Index	# Relatives with Br.Can
Obs	9,197	9,197	8,759	8,759	9,197	9,197
Pseudo R-sq	0.1373	0.1383	0.0414	0.0689	0.1082	0.1086

1) Reported are coefficients from a Tobit for the number of mammograms during the last 6 years, an ordered Probit for the Breast Cancer Risk variable and a Probit for the doctor recommendation variable. The sample consists of 2005 women aged 30-85, since the doctor recommendation variable is only available for 2005. Standard errors are reported in parentheses. For the interactions: ** indicates that we reject one-sided tests at a significance level of 1% , * at the 5% and + at the 10% level. For the other coefficients these symbols refer to the usual two-sided tests. All specifications control for a full set of age, health care coverage, income bracket, and education indicators and also contains a risk-factor main effect.