DISCUSSION PAPER SERIES

IZA DP No. 3548

Education, Information, and Improved Health: Evidence from Breast Cancer Screening

Keith Chen
Fabian Lange

June 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Education, Information, and Improved Health: Evidence from Breast Cancer Screening

Keith Chen
Yale University

Fabian Lange
Yale University
and IZA

Discussion Paper No. 3548
June 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3548
June 2008

ABSTRACT

Education, Information, and Improved Health: Evidence from Breast Cancer Screening

While it is well known that education strongly predicts health, less is known as to why. One reason might be that education improves health-care decision making. In this paper we attempt to disentangle improved decision making from other effects of education, and to quantify how large an impact it has on both a patient's demand for health services, and that demand's sensitivity to objective risk factors. We do this by estimating a simple structural model of information acquisition and health decisions for data on women's self-reported breast-cancer risk and screening behavior. This allows us to separately identify differences in the ability to process health information and differences in overall demand for health. Our results suggest that the observed education gradient in screening stems from a higher willingness-to-pay for health among the educated, but that the main reason why the educated respond more to risk factors in their screening decision is because they are much better informed about the risk factors they face.

JEL Classification: I10, I12, I20, D83

Keywords: education, allocative efficiency, health

Corresponding author:

Fabian Lange
Department of Economics
Yale University
37 Hillhouse Ave.
New Haven, CT 06520
USA
E-mail: fabian.lange@yale.edu

1 Introduction:

The fact that education and health are extremely correlated represents one of the fundamental facts in health economics. This link between education and health is so close that Grossman and Kaestner (1997) suggest that education may be the most important socioeconomic correlate of good health. A number of explanations have been offered to explain this tight correlation. Education raises incomes and consequently expenditures on health as long as health is a normal good. In turn, healthy individuals can expect longer life-spans and thus can expect to benefit more from investing into education. Furthermore, health and education will be affected in similar ways as the discount rate varies, because health and education both represent investments into the future. Grossman's (1972) seminal model of health capital shows how education and health are linked for all of these reasons, but also hypothesizes that education may directly enable individuals to produce health capital more effectively. That is, might education directly improve the efficiency of health investments by improving an individual's health-care decision making?

Plausibly, the educated are more efficient at investing into health because they know more about health, and hold an advantage in processing health information. If the educated process health information more effectively, they would allocate health resources more effectively and thus produce more health from the same amount of resources. This hypothesis is known as the *allocative efficiency hypothesis* of education in health investments.

To date however, the empirical evidence in support of the allocative efficiency hypothesis is largelylimited by a fundamental identification problem; Do more highly educated individuals make different health-related choices because they value health differently or because they differ in their knowledge about the consequences of their choices.

Two interesting studies by de Walque (2004a, 2004b) illustrate this difficulty in the context of two very different environments. In both of these studies, de Walque (2004a, 2004b) finds that new information on health induces faster and more pronounced responses among the educated. De

Walque (2004a) finds that as new information about smoking risks became available in the 1950s and 1960s, the more educated quit smoking at higher rates than did the less educated. De Walque (2004b) documents a similar pattern in sexual behaviors in Uganda as the HIV epidemic unfolded.[1] These findings are consistent the notion that more educated individuals process new medical information more efficiently and adjust more rapidly to this new information. However, the same patterns in behavior are consistent with the hypothesis that no differences in health knowledge exist between the educated and uneducated, but that the educated respond more to the new information because they have more elastic demand for health. Indeed, the greater drop in smoking rates observed in the US since the 1950s and the more pronounced decline in risky sexual behaviors among the educated in Uganda are consistent with nearly every theory which links education and health.

More direct evidence for the allocative efficiency hypothesis comes from Kenkel (1991a, 1991b). Kenkel showed that when surveyed, more educated individuals are better informed about the health risks associated with smoking, drinking and lack of exercise. He also finds that those who are aware of the specific health risks associated with these risk factors are more likely to abstain from smoking, drinking, and to exercise. To our knowledge Kenkel (1991a, 1991b) was the first to show that health knowledge correlates both with education and with health behaviors. As such, these papers provide the closest direct confirmation of the allocative efficiency hypothesis.[2] However, these papers do not provide information on whether the allocative efficiency hypothesis is quantitatively important relative to education-driven increases in the demand for health. That is, current studdies do not let us judge adequately whether differences in the ability to process health information have large quantitative implications for health care decisions.

[1] Prior to the spread of AIDS, educated men in Uganda tended to have more sexual partners than those with less education. However, by the early 1990s, educated men were among the first to adjust their sexual behavior and reduce the number of sexual partners as well as the incidence of unprotected sex.

[2] Kenkel (2000) provides an excellent survey of the empirical evidence for the allocative efficiency hypothesis.

At the same time, it has become increasingly relevant to understand whether constraints in the ability to process health information among the less educated are large. 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 many bewildering options. If the skills required to negotiate the modern health care system are partially aquired 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 is increasingly unequally distributed (see Cutler, Meara, and Richards (2008)). If policy makers want to expand the benefits of modern medical care to the less educated, they need to understand whether the less educated achieve worse heath outcomes because they can't afford health care, or because they lack the relevant health-care knowledge and decision making skills.

Furthermore, health care expenditures now comprise an unprecedented share of GDP. At the same time relatively cheap ways of improving health and longevity seem underutilized; few individuals receive flu shots, despite the fact that more than 200,000 people are hospitalized for flu complications a year, and around 36,000 of those die (CDC http://www.cdc.gov/flu/keyfacts.htm). Cutler & Lleras-Muney (2007) estimate that each additional year of schooling increases the probability of receiving a flu shot by 1.7 percentage points, suggesting that education may be key to increasing the utilization of cost-effective preventative care and screening.

In this paper, we use data from Cancer Screening and from subjective statements about cancer risks to provide new qualitative and quantitative evidence on the allocative efficiency hypothesis. In particular, we examine whether the educated are better informed about their individual risks for developing cancer. We start from the observation that those who are better informed about their risk will be more responsive to the presence of risk factors *both* in their self-reported risk status and in their demand for preventative health care. Simply observing the variation in the demand for health with risk factors will however not be sufficient to distinguish between

allocative efficiency and other competing explanations for the link between education and heatlh. In Section 2, we discuss this identification probelm more formally.

We then present (Section 3) reduced form evidence that educated women are more responsive to the presence of risk factors when reporting their subjective risk of developing breast cancer.[3] We also show that educated women respond more to risk factors when they decide whether to be screened for cancer. Together these two findings suggest that education enhances allocative efficiency. The finding that risk factors also affect screening more among the educated documents the relevance for behavior. However, the reduced form evidence does not enable us to quantify directly how large a role differences in the ability to process health information by education play in generating differences in health behavior. If we want to quantify the role of education in enhancing allocative efficiency, then we need to specify and estimate a model of health information and decisions. We specify a simple model of health information and demand in Section 4. Our model allows for both differences in the ability to process information and for differences in the willingness-to-pay for health improvements across education levels. In Section 5, we estimate both the demand parameters and the parameters of the information model using data on subjective risk assessments, objective risk factors and cancer screening decisions.

Our estimates suggest that allocative efficiency indeed plays an important role in generating differences in screening behavior across education groups. We find that almost the entire differences in the responsiveness of cancer screening decisions across education groups are driven by differences in the ability to process information. Yet, this finding should not be misconstrued to mean that the observed health gradients are entirely due to differences in allocative efficiency. Indeed, the estimate from our model imply that the observed differences in the ability to process information ac-

[3] We pay particular attention to breast cancer because there exists a well developed medical risk model for breast cancer screening. However, we also show results from other cancers and in particular colorectal cancer that are consistent with our reduced form findings from breast cancer.

count for very little of the observed overall gradient in screening. Instead, the overall gradient in screening with education is primarily driven by differences in the demand for health across education groups. Overall, our results suggest that the impact of differences in allocative efficiency on the overall health gradient depend on the particularities of the health decision considered. For breast cancer screening, we find that allocative efficiency can not explain why less educated screen at much lower rates than the more educated. However, our results also imply that differences in the ability to process information do indeed play a quantitatively important role in determining observed health behaviors.

2 The Identification Problem: Separating Variation in Subjective Mortality Risk from the Demand for Health

Ultimately, empirical research needs to strive to both test and quantify how important the allocative efficiency hypothesis is for generating observed differences in health behaviors across education. Empirical researchers face the difficulty to separately identify how important information and demand based explanations are; this problem can not be overcome based only on observed differences in health behaviors across education. Instead, we require additional information that allows us to either identify the demand for health or the distribution of information in the population from another source. In this section, we strive to clarify the problem faced by empirical researchers.

To build intuition, consider how smoking trends have varied with education over the last half century and whether we can learn about differences in the ability to process health information from these differential trends in smoking across education. During the 1950s and early 1960s, evidence based health research revealed the considerable health risks associated with smoking, and numerous public health campaigns sought to inform the public of these risks. As demonstrated by De Walque (2004) using the NHIS,

over that period individuals with higher levels of education quit smoking at higher rates. Prior to 1950, education and smoking were positively correlated, but by 1960 the less educated were smoking at higher rates than the more educated.

The fundamental problem we face is to determine why the educated responded more rapidly to the new information. There are several reasons why the gradient between education and smoking reversed. On the demand side, the educated may naturally place a higher monetary value on survival and hence face greater incentives to quit smoking once smoking is commonly known to be harmful. Similarly, those with higher discount rates may have both valued education and health more highly. Consequently, those with higher discount rates may have both invested more in education and quit smoking at higher rates as new information on the health effects of smoking became available. Alternatively, education may have been positively correlated with either the exposure to or the more rapid absorption of new health research. This information effect would likewise induce the educated to quit smoking at higher rates than the less educated. The observed implications of the information and the demand effect on smoking rates is the same.

To illustrate this identification problem formally, we start from a simple model of health decisions, and show why observational data on health decisions alone does not reveal whether education improves allocative efficiency. Individuals make a mortality relevant and costly dichotomous decision represented by $\theta \in \{0, 1\}$. Each alternative is associated with a financial cost $\{c_{0,i}, c_{1,i}\}$ and a subjective survival probability $\{\widetilde{\pi}_{0,i}, \widetilde{\pi}_{1,i}\}$. Subjective survival probabilities are denoted with tildas to differentiate them from objective survival probabilities $\{\pi_{0,i}, \pi_{1,i}\}$.The costs and also the subjective (and objective) survival probabilities attached to each alternative vary across individuals. Let $\Delta c_i = c_{1,i} - c_{0,i}$ and $\Delta\widetilde{\pi}_i = \widetilde{\pi}_{1,i} - \widetilde{\pi}_{0,i}$represent the difference in costs and the gain in survival associated with each alternative and write $\Delta c_i = \Delta c + \varepsilon_i$.

The decision can then be represented as:

$$\theta = 1 \Leftrightarrow VSL\Delta\widetilde{\pi}_i - \Delta c \geq \varepsilon_i \tag{1}$$

Here the parameter VSL represents the marginal rate of substitution between survival and consumption. This parameter is known as the Value of a Statistical Life. For simplicity, I will assume that the individual heterogeneity in costs ε_i is independent of all other variables. This heterogeneity is distributed with F_ε.

The typical data-set will contain data on health behaviors and sometimes information on the objective risks that individuals face. We also assume that education e and potentially other relevant control variables x are observed. To simplify the problem even further, assume that the demand parameters $(VSL, \Delta c)$ vary with education and x, but not with the objective risk π. Thus, let $(VSL(e,x), \Delta c(e,x))$ represent the Value of a Statistical life and the average cost of the health decision conditional on the education and the variables x observable in the data.

The data therefore allows us to identify

$$\Pr(\theta = 1|x, e, \Delta\pi) = \int (1 - F_\varepsilon(VSL(e,x)\tau - \Delta c(e,x))) g(\tau|\Delta\pi, e, x) d\tau \tag{2}$$

Here, $g(.|\Delta\pi, e, x)$ represents the density of subjective survival gains $\Delta\widetilde{\pi}$ conditional on the objective gains $\Delta\pi$ as well as (e,x). Clearly, any observed differences in health decisions with education can be reconciled with subjective beliefs about survival gains that are common across education $g(.|\Delta\pi, e, x) = g(.|\Delta\pi, x)$ by appropriate choice of the demand parameters $\{VSL(e,x), \Delta c(e,x)\}$. Similarly, by appropriately choosing $g(.|\Delta\pi, e, x)$ we can reconcile any observed patterns in health choices across observable characteristic with arbitrary values of $\{VSL(e,x), \Delta c(e,x)\}$.

Consider the findings of De Walque (2004a) in this light. He showed that smoking rates among those with high levels of education declined more rapidly as it became known that smoking was dangerous. This finding is of course consistent with the assumption that $g(.|\Delta\pi, e, x)$ changed more rapidly among the more educated. Since the more educated updated their subjective risk assessments about the dangers of smoking more rapidly as the new information on health risks became available, smoking rates among the more educated started to decline earlier and more rapidly among them. The

allocative efficiency hypothesis therefore is consistent with the observed facts about smoking in the last 50 years. However, these facts are also consistent with the hypothesis that both the educated and the less educated learned about smoking risks at the same time and that the more educated simply reacted more strongly to this information because they attached higher valuations ($VSL\left(e,x\right)$) to survival. The educated quit at higher rates as new information about the health risks of smoking became available, but not because they were more aware of the harmful effects of smoking. The observed trends in smoking thus illustrate the problem of identifying the role of information processing and demand differences in health decisions using observational data only.

A similar problem arises if we want to exploit observed variation in how screening rates vary with risk factors and education. The differences in how responsive the educated are to variation in objective risks might be due to variation in the pdf of subjective beliefs $g\left(\Delta\widetilde{\pi}|\Delta\pi,e,x\right)$ or due to variation in the demand parameters $(VSL(e,x),\Delta c\left(e,x\right))$. We can not separately identify the information and the demand contribution to the differences in screening behavior without imposing further restrictions.

One possibility to make progress is to assume that we have information about the demand parameters $(VSL(e,x),\Delta c\left(e,x\right))$ from other sources. For instance, we might specify how the demand for health is related to consumption and income (e.g. Murphy and Topel (2007)) and then use data on consumption and incomes across education as well as the costs of screening to link and subject's VSL and education. This approach requires a lot of information on health demands. Empirical estimates on the Value of a Statistical Life are very imprecise and even less is known on how the VSL varies with education. Furthermore, it is often very difficult to estimate the full costs of health decisions. In the case of the smoking example we would need to measure the value attached to smoking by smokers. For the screening example below, the full costs of screening for cancer include costs that arise because of physical and psychological pain and suffering associated with screening and with various diagnostic procedures associated with screening. Furthermore, if the health decisions are dichotomous decisions,

then we typically also need the variance of the cost $\Delta c + \varepsilon_i$ across individuals. The data at our disposal has relatively little information about the costs of screening, nor about consumption or incomes of individuals.

An alternative is to try to measure the subjective risk. It is this avenue that we exploit in this paper. The reason why we are able to make progress along these line is because the National Health and Interview Survey (NHIS) of 2005 has collected unique data on both objective and subjective risks. We specify an information model that allows us to estimate the joint distribution $G\left(\Delta\pi, \Delta\widetilde{\pi}|e, x\right)$ of subjective and objective risk conditional education e and x from subjective risk statements provided by respondents to the NHIS and measures of their objective risk. Based on this estimated distribution and on the observed variation in screening behavior, we can then estimate the demand parameters $\left(VSL\left(e, x\right), \Delta c\left(e, x\right)\right)$. The estimated parameters of the information model and the demand parameters together allow us to deduce how much of the variation in screening behavior with objective risk, education and other variables is attributable to differences in the way educated and less educated individuals process information and how much of it can be attributed to demand differences.

Before we turn to this structural estimation exercise, we will however first provide a detailed description of the data including an extensive reduced form analysis of the data.

3 Reduced Form Evidence from Cancer Screening and Subjective Cancer Risk

In the previous Section, that observational data on health behavior alone does not suffice to test or quantify the allocative efficiency hypothesis. In this Section, we introduce reduced form evidence from both cancer screening behavior and from subjective risk assessments related to cancer to provide qualitative evidence on the allocative efficiency hypothesis.

We examine whether education changes the way individuals respond to the presence of risk factors when they assess their subjective cancer risk

and also when they decide to screen for cancer. We show that as objective risk of developing cancer increases, educated individuals are differentially more likely to report that they are of high cancer risk. Furthermore, risk factors are also differentially more likely to induce educated rather than uneducated individuals to undergo cancer screening. Both these empirical facts are consistent with the allocative efficiency hypothesis that education raises awareness of risk factors and thus enables individuals to make better health decisions.

3.1 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, objective cancer risks and cancer screening made available through the National Health and Interview Survey (NHIS). The NHIS collected individual assessments of subjective cancer risks for Breast Cancer and for Colorectal Cancer in its cancer control module in 2005. In addition, the NHIS cancer control modules recorded screening choices made by individuals and also recorde a number of variables that allow us to implement a well known model of breast cancer risk prominent in the medical literature. We do not have a similar risk model for the other cancers in the data. For that reason our analysis focuses on breast cancer.

3.1.1 The Gail Index

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 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 1]

Table 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.[4]

3.1.2 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

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

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. We are limiting ourselves to non-Hispanic sample adult females. Only women aged 30 and older were asked questions relating to breast cancer screening. 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 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.

For the screening decision, our dependent variables are an indicator vararіable 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.[5]

Table 2 presents summary statistics of the variables used in the analysis.

[Table 2]

The distribution of socioeconomic variables is typical for the population of women aged 30-85. The average woman has more than a high school degree and the average years of schooling completed increases between 2000 and 2005, reflecting the higher rates of education among women born in later cohorts. Less than 10% of women in this sample lack health care coverage of any kind. This is partially because a sizeable fraction of women is older than 65 and consequently covered by Medicare. In 2005, about half of all women reported that they were recommended by a medical professional to be screened for breast cancer.

Screening rates are high and increased slightly between 2000 and 2005. The average Gail Index in the population is slightly greater than 1. It increases between 2000 and 2005. This is driven by the observed decline in age at menarche, the increase in the number of individuals with direct family members that had breast cancer and the increase in age at first birth. There is quite a bit of variation in the Gail Index in the population. The standard deviation is about 0.8 in both years.

The subjective risk assessment variable differs between 2000 and 2005. In 2000 respondents were asked whether their risk of developing cancer overall was low, medium or high, whereas in 2005 they were asked whether the were less likely, about as likely or more likely than the average woman to develop breast cancer specifically. In 2005, about 12% of women overall report that they perceive themselves to be more likely to develop breast cancer, whereas about 35% believe themselves to be less likely to develop breast cancer.

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

3.2 Reduced Form Specification

For the remainder of Section 3, we examine how cancer risk correlates with a variety of outcomes. The most important outcomes we consider are the subjective breast cancer risk assessments elicited from respondents to the NHIS and information on the breast cancer screening behavior of women. We also consider information on the screening recommendations made to women by medical professionals and information on screening and subjective risk related to other cancers. Most of the outcome variables we consider are either binomial or multinomial categorical variables.

We therefore consider reduced form specifications that account for the categorical nature of the variables. These specifications are usually based around an index formulation, where the index is of the following form:

$$Ind_i = \beta_x' x + \beta_G * G + \beta_s * S + \beta_{G \times S} * S * G + \varepsilon_i \tag{3}$$

Here the variable S measures education and the variable G represents the Gail index. The residual ε_i is assumed to be normal. We include a vector of controls x, which are occasionally themselves interacted with the Gail index. Generally we control for age using a full set of age dummies. We have also examined whether all our results are robust to limiting the analysis to women younger than age 60. Our specifications also typically control for income using the categorical income variables available in the data.[6] We examine all our results for robustness to relaxing the linear schooling effects to allow for non-linear specifications of the education main effects as well as for dropping the income measures. In the tables we report various results from additional robustness checks. Among these are specifications that control for various measures of health care coverage. Unless we found deviations from our main findings, we do not discuss the results from the various robustness checks in the text.

[6] The income variable is a categorical variable that reports family income as a multiple of the poverty line. The highest category represents incomes 5 times above the poverty line. There is a sizable number of invalid answers for this question and we include a dummy for missing answers.

The index then determines the observed variable θ_i. Depending on the variable of interest, θ_i will either be a multinomial ordered categorical variable (for the subjective risk assessments) or a binomial categorical variable (for the main screening specification and when we consider whether a doctor recommended screening). We also report estimates from a screening model where the dependent variable is the number of screens a woman underwent during the last 6 years. In that case, θ_i will be a left censored random variable with $\theta_i = Ind_i * 1\left(Ind_i \geq 0\right)$.

3.3 Subjective Risk Assessments and Education

The first empirical question we consider is whether educated women are more likely than less educated women to report themselves to face a high risk of developing breast cancer when they indeed face a high objective risk of developing breast cancer risk. For this purpose, we estimate ordered probit models based on specification (3) using the 2000 and 2005 subjective cancer risk assessments.

If women with higher education do more accurately predict their individual cancer risk, then objective risk factors (as measured by the Gail Index) should be more strongly correlated with subjective risk assessments among educated women. The results of estimating ordered probits for the risk assessment confirm this hypothesis for both the 2000 and 2005 data.

[Table 3]

We are interested in the interaction between schooling and Gail Index. The positive interaction reported for 2000 and 2005 (col. 1 and 4) confirms that self-reports of subjective cancer risk correlate more strongly with objective risk among educated women.

In 2000, the subjective cancer risk question referred to the risk of cancer generally and in 2005 to breast cancer specifically. We therefore expect the link between the Gail index to be stronger in 2005 than in 2000; we find

that the main effects and the interactions with education are stronger in 2005 than in 2000. We will estimate the structural information model in Section 4 and 5 using the questions from 2005 only.

The interaction of education with the Gail model survives including income measures and interacting these with the Gail index. The interaction of the income variable with the Gail index is not significant (individual or jointly) in 2000 and close to marginally significant (with p-values around 0.11) in 2005. The sign of these Income $\times$ Gail interactions is also positive and thus indicates that those with higher incomes are more sensitive in their risk assessments to the presence of risk factors. We discuss the Income $\times$ Gail interactions in more detail below.

Family risk plays an important role in driving the association between the Gail index and subjective risk. Once we include family risk dummies (col. 3 and 6) and interact family risk with schooling, then the main effects of the Gail index and the interactions of the Gail with schooling vanish. Family risk seems to be more salient than other risk factors for determining subjective risk.

Overall, we find that indeed the educated respond more to the presence of risk factors than the less educated when answering questions about personal breast cancer risk. Much of this is driven by family risk, but it indicates that indeed the more educated have an advantage in processing risk relevant information.

3.4 Cancer Screening and Education

We next show that the educated also tend to respond more to risk factors when they make cancer screening decisions. We estimate specification (3) for a Probit model where the dependent variable θ indicates whether a woman has ever received a mammogram. A positive estimate of $\beta_{G\times S}$ indicates that educated women are relatively more likely to receive screens if they face objective cancer risks compared to less educated women.

As an alternative dependent variables we also consider the number of screens a woman has received during the last 6 years. Thus, we also estimate

a Tobit on specification (3) where the dependent variable θ_6 represents the number of mammograms during the last 6 years.

Table 4 Panel 1 presents the parameter estimates for both specifications obtained on the 2000 data. In columns 1 and 6 (our baseline) the control set includes a set of dummies for the income variable. The baseline specification also includes a full set of dummies for age and race. Table 4, panel 2 reports the analogous results for the 2005 data and in table 5, panels 1 and 2 we show the marginal effects of the Probit specification to provide an indication of the magnitude of these effects.

[Tables 4 and 5]

We find (and this finding is robust across specifications), that more educated individuals respond more to risk factors than do less educated individuals. The effect on the propensity to screen (the marginal effect) reported in table 5, column 1 gives an indication of the size of the effect. In 2000, an increase in the Gail Index by one unit (doubling the relative risk of breast cancer) raises the probability of screening by about 3 percentage points more among college educated women than among women with only a high school degree. Given the overall rate of screening in the population of about 75%, this represents a sizeable additional responsiveness. The findings are similar when we consider the intensity of screening using the Tobit specification.

Interestingly, income and the Gail index also interact in the screening decision (cols 4 and 9). There are various interpretations for this fact. One possibility is that demand factors of the type discussed in Section 2 generate a higher response of screening to risk factors among those with high incomes. Another interpretation of this interaction is that income/earnings and education are both measures of human capital. This explanation has the virtue of being consistent with the finding reported in table 3 that subjective income and the Gail model do interact in the subjective risk assessment questions.

3.5 Better Doctors?

The results reported in tables 2 to 5 show that educated women respond more strongly to the presence of objective risk factors when they report their subjective risk assessments and also when they decide on screening. This positive interaction both in the subjective risk data as well as in the screening decision is suggestive that the educated are indeed better at processing health information. An alternative explanation for this fact is however that the educated or those with more income might simply have access to better doctors who help them make this decision.

In table 6 we present 3 different ways of testing for this hypothesis using a variable from the 2005 NHIS that reports whether a doctor has recommended a mammogram within the last 12 months. In columns 1 and 3 we show that women who were recommended to receive a mammogram were indeed more likely to undergo a screen and they also perceived themselves to be at higher risk. However, there is no evidence (column 2) that the doctor recommendations interacted with either education or with the Gail index or with the triple-interaction in generating concerns about cancer screening. Furthermore, we can not find a statistically perceptible change in the Schooling $\times$ Gail interaction when we control for the doctor recommendation alone or including its interactions with the Gail and with education. Our results are similar for the screening decision. If anything both the interaction between the Gail and education becomes slightly stronger both in the subjective assessments and in the screening decision. These results indicate that doctors recommendation did not disproportionately induce educated, high risk women to receive screens.

In columns 5 and 6 we directly consider the question whether schooling and the Gail index interacted in determining whether a doctor recommended screening. There is no evidence that doctors responded more to the presence of risk factors when advising patients with high rather than low education levels. Overall, we do not find evidence that educated women visit doctors that are more sensitive to risk factors when recommending screening.

[Table 6]

3.6 Colorectal and Cervical Cancer Screening

Finally, we consider screening behavior related to colorectal and cervical cancer. We do not have access to a risk model comparable to the Gail model for these cancers and therefore limit ourselves to considering how screening behavior and subjective risk assessments for these cancers vary with the occurrence of these cancers in the family. We will refer to family risk below as a variable that indicates whether a cancer of the relevant type has been observed among direct relatives of the respond.

Table 7 summarizes how family risk interacts with education in both the screening decisions and subjective risk assessments for Breast Cancer, Colorectal Cancer and Cervical cancer. Each column corresponds to a different screening decision or subjective risk assessment. Reported are only the interactions with family risk. Each specification includes both the family risk for the cancer in question as well as the family risk for the cancers that are not relevant for the test or the subjective risk assessment. The estimnates are obtained by estimating specification (3) with full sets of dummies for age, education, ethnicity, and income.

[Table 7]

In columns (1)-(4) we examine the decision to screen for colorectal, cervical and breast cancer. We pool the data across both 2000 and 2005 and fully interact all controls with year dummies. Colonoscopies and Home Blood Stool tests screen colorectal cancer and pap smears screen for cervical cancer. If the educated are more responsive to the presence of specific cancer risk, then we expect the reported interaction on family breast cancer risk in column (1), family colon cancer risk in columns (2) and (3) and family cervical cancer risk in column 4 to be positive. We do not expect interactions of either sign with any of the cancers that are not subject of the screen or subjective risk question.

In addition to the subjective risk question about Breast Cancer, the NHIS 2005 also includes a similar question for colorectal cancer. We therefore estimate ordered probits with interactions between various cancer types in the family and education to examine whether again the subjective risk assessments of educated individuals are more responsive to the presence of risk factors than are those of the less educated. We report these estimates in columns (5) and (6).

The only statistically or economically significant coefficients in table 7 are those on the specific cancers in question. The coefficients on colorectal family risk for the blood stool test is economically irrelevant. Home blood stool tests are known to be of low predictive power, potentially explaining that we find little systematic behavior with respect to these. The coefficients on the pap smear test is statistically not significant, the point estimate is however sizeable and of the expected sign. For subjective risk we observe that the educated respond more than the less educated to the presence of breast cancer risk among direct relatives when they judge their own breast cancer risk and the same is true for colorectal cancer risk. They don't generally seem more responsive to the presence of cancer among other family members.

Overall, the results reported in table 7 are broadly consistent with the hypothesis that the educated both respond more to the presence of risk factors when they decide to be screened for cancer as well as when they report their subjective risk of developing cancer.

3.7 Discussion of Reduced Form Results

Overall the detailed examination of the data revealed the robust result increases the responsiveness to the presence of cancer risk factors both in subjective risk assessments and in the cancer screening decisions. Our evidence resembles the type of evidence provided by Kenkel (1991a, 1991b) in that we examine both survey responses related to health knowledge and observational data from behaviors in reduced form. Because we examine a different type of health behavior and health risk, they provide additional

support for the allocative efficiency hypothesis.

What we can not do is determine the extent to which information rather than demand effects drive the interaction of education with risk factors using only the screening decision. For that purpose one does need to combine information about subjective risk assessments of individuals with observed screening decisions. This requires an explicit model of health information.

4 A model of objective breast cancer risks, subjective risk assessments and screening decisions

The NHIS contains data on the subjective risk of developing breast cancer, on objective risk factors and on the decision to screen. The challenge we face is to write down an explicit, estimable model that links these components. In particular, we need to model how agents form subjective risk judgements and how the breast cancer screening decision is related to these subjective risk judgements. Clearly, this model does not realistically describe the mental processes that take place when individuals assess risks, but it is intended to describe the joint distribution of subjective and objective risks as well as the screening decision in a transparent manner.

4.1 The information model

According to the medical literature, the objective risk factors facing an individual can be summarized by an Index G_i called the Gail Index. For our purposes we can think of the Gail Index as the 5 year probability of developing breast cancer.[7] We construct the G_i for each individual in the data using the information in the NHIS.

Our risk model centers around a latent variable t_i describing a womans cancer status. If $t_i < 0$, then the woman will develop breast cancer within the next 5 years.

[7] The Gail Index is actually a measure of the relative risk of breast cancer compared to women of the same age. We translate this relative risk into actual probabilities using the base-line probability of developing breast cancer.

Next we associate the Gail Index G_i with a variable g_i that lives in the same space as the latent risk variable t_i. We relate the G_i and g_i to t_i using assumption A1:

Assumption A1 The latent variable t_i is distributed according to $t_i = g_i + \varepsilon_t$ with $\varepsilon_t \sim N(0,1)$.

Because G_i represents the probability of developing breast cancer we have $P(t < 0|g_i) = \Phi(-g_i) = G_i$ where $\Phi(\cdot)$ is the normal cdf. This defines a value g_i associated with each G_i.[8]

Assumption A2 Individuals learn about their risk by drawing a signal $s_i = g_i + \varepsilon_s$ where $\varepsilon_s \sim N\left(0, \sigma_s^2\right)$ and $\varepsilon_s \perp \varepsilon_t$.

This simple structure defines $F(s|g) = \Phi\left(\frac{s-g}{\sigma_s}\right)$, the distribution of the signal conditional on the Gail index. The Gail index is a discrete random variable that is observed in the data and defines a marginal distribution $P(g)$. This therefore delivers the joint distribution $F(s,g) = \Sigma_g P(g) * F(s|g) = \Sigma_g P(g) * \Phi\left(\frac{s-g}{\sigma_s}\right)$.

Next, we need to describe how signals s are related to the subjective risk assessments available in the data. Our information about subjective risks comes in the form of answers to: "Do you believe that you are of low, medium, high risk of developing breast cancer?". Assumption A.3 provides the link between the answer to this question and the signal s.

Assumption A3 There are two cutoffs (s_H, s_L) such that if $s > s_L$ then the woman declares herself to be of low risk and if $s < s_H$ then she will say that she is of high risk. In between, she calls herself of medium risk.

To estimate the variance of the signal σ_s^2 we will use the joint distribution of the answers to the subjective risk questions with the objective risks faced by individuals. Conditional on the Gail index, 3 parameters in the model determine whether an individual declares herself to be of low or high risk. These are $\left(s_L, s_H, \sigma_s^2\right)$ and represent the parameters of the model that need to be estimated using the observed fractions of individuals who declare themselves to be of low or high risk conditional on the Gail Index. Denote

[8] Setting the variance of the residual ε_y to unity represents a normalization as long as the mapping from G_i to g_i is unrestricted.

these fractions by $\left\{\widehat{P}_{L,g}, \widehat{P}_{H,g}\right\}$. Then, the information model implies that

$$\begin{aligned} \widehat{P}_{L,g} &= 1 - \Phi\left(\frac{s_L - g}{\sigma_s}\right) \\ \widehat{P}_{H,g} &= \Phi\left(\frac{s_H - g}{\sigma_s}\right) \end{aligned} \tag{4}$$

These moment conditions relate the Gail index to the subjective risk factors and allow us to estimate the parameters of the information model. Note that we assume that individuals do know the average breast cancer risk conditional on age. Within age, we construct the moments (4) in the manner described above.

4.2 The decision model

Screening lowers the mortality risk associated with breast cancer through early diagnosis. Let d_θ represents the mortality reduction due to early diagnosis that can be achieved by screening. Let $c_i = c + \varepsilon_c$ represent the costs of screening which depends on a common component c and an individual component ε_c. This cost includes both financial costs and psychic costs associated with the screening procedure.

A woman with signal s_i therefore decides to screen ($\theta = 1$) iff:

$$VSL * d_\theta * P(t_i < 0|s_i) \geqslant c + \varepsilon_c \tag{5}$$

We impose that $\varepsilon_c \sim N(0, \sigma_c)$. Conditional on the private signal s_i, a fraction $\Phi\left(\frac{V}{\sigma_c} * P(t_i < 0|s_i) - \frac{c}{\sigma_c}\right)$ will decide to screen where $V = VSL * d_\theta$.

Once we estimated the parameters of the information model we can construct the joint distributions of (s, g, t) and therefore also $P(t < 0|s)$.[9] We can also use estimate of $F(s, G)$ and the estimated cutoffs (s_L, s_H) to

[9] Assumption A.1 implies that $P(t < 0|g_i) = \Phi(-g_i)$ and using P(g) we get the prior distribution $P(t < 0) = \Sigma_g \Phi(-g_i) P(g_i)$. The posterior distribution after receiving the individual signal s is $P(t < 0|s) = \Sigma_g \Phi(-g_i) * P(g|s)$. Based on the estimated σ_s^2 from the information model we can construct $F(s, g)$ and can therefore also obtain $P(g|s) = \frac{F(s|g) * P(g)}{f(s)}$. Therefore, for each signal we can generate $P(t < 0|s)$ after estimating the information model.

determine the fraction of women screening for breast cancer conditional on G and conditional on reporting herself to be of low risk as:

$$P(\theta = 1|G, s > s_L) = \int_{s>s_L} \Phi\left(\frac{V}{\sigma_c} * P(t_i < 0|s_i) - \frac{c}{\sigma_c}\right) dF(s|g) \quad (6)$$

Similar moments are obtained for medium and high subjective risk and these moments allow us to estimate the parameters $\left(\frac{V}{\sigma_c}, \frac{c}{\sigma_s}\right)$ using minimum distance.

There is no need to estimate the full vector $\{VSL, d_\theta\}$. Instead we estimate $\frac{V}{\sigma_c} = \frac{VSL*d_\theta}{\sigma_s}$ and $\frac{c}{\sigma_s}$. These are sufficient to generate the screening demands conditional on the subjective and objective risks.

5 Estimating the Structural Model

5.1 The Information model

As described above, the parameters of the information model can be estimated using the moment conditions in equation (4). To estimate the information parameters we allow for separate cutoffs (s_L, s_H) by education and 5-year age-group. We restrict σ_s to be constant across age, but estimate σ_s separately for each of the three education levels $(\leq 12, 14, \geq 16)$.

Table 8.shows our estimates of σ_s for each education level. In panel A we show the results for the full risk model. The reduced form results in tables 3 however suggested that the presence of cancer in the family is more salient than the other risk factors entering the Gail index. We therefore also estimated the information by constructing the Gail index using only the information on the family risk. The results when estimating this model are presented in panel B. In our discussion we will largely limit ourselves to the full risk model.

We boot-strap the measures of dispersion.[10] For the full risk model and some of the boot-strapped samples, the relation between the Gail and the

[10] It takes significantly longer to estimate the full risk model rather than the model using family risk only. For this reason we currently bootstrap the full risk model 60 times and the family risk model 500 times.

subjective risk assessment is negative. For these samples, the estimates of σ_s go to infinity. We therefore report three measures of dispersion. First, we report the standard deviation of $\widehat{\sigma}_s$ and the 90% confidence intervals conditional on having a finite $\widehat{\sigma}_s$ and then we show the number of samples (out of 60) for which $\widehat{\sigma}_s \to \infty$.

[Table 8]

These estimates confirm the finding from the multinomial Probit model estimated in Section 2 that less educated women are less informed about the risks they face. This is true if we compare the signal noise σ_s of women with 12 years of education with that of women with more education, but also if we compare the signal noise of women with 13-15 years of completed schooling relative to that of women with 16 or more years of education. Overall, we find a much stronger relation between the true risk and the individual subjective risk among educated than less educated women.

5.2 The Decision Model.

The estimated information model describes how the private signal s and the objective Gail are linked and consequently describes how the subjective morality risk is distributed conditional on the Gail signal and the reported subjective risk assessment. Based on this distribution we can estimate the parameters of the decision model. In particular, we can numerically generate the predicted probability of screening conditional on the information model and the parameters of the decision model. The information model together with the decision model therefore generates the following moments for each point of the support of the Gail distribution:

$$\begin{aligned} P\left(\theta=1|g,s>s_L\right) &= \int \Phi\left(\frac{V}{\sigma_c} * P(t_i<0|s_i) - \frac{c}{\sigma_c}\right) dF\left(s|g,s>s_L\right) \\ P\left(\theta=1|g,s_H<s<s_L\right) &= \int \Phi\left(\frac{V}{\sigma_c} * P(t_i<0|s_i) - \frac{c}{\sigma_c}\right) dF\left(s|g,s_H<s<s_L\right) \\ P\left(\theta=1|g,s<s_H\right) &= \int \Phi\left(\frac{V}{\sigma_c} * P(t_i<0|s_i) - \frac{c}{\sigma_c}\right) dF\left(s|g,s<s_H\right) \end{aligned} \tag{17}$$

The empirical counterparts to these moments are the observed fractions of individuals that are screened for breast cancer conditional on their objective cancer risk and their subjectively reported cancer risk.

We estimate the demand parameters $\left(\frac{V}{\sigma_c}, \frac{c}{\sigma_c}\right)$ separately for each education level using minimum distance and the moments in equation (7). In order to allow the Value of a Statistical Life to vary with age, we specify $\left(\frac{V}{\sigma_c}, \frac{c}{\sigma_c}\right)$ as fourth-order polynomials in age and consequently need to estimate a total of 8 parameters for each education group. The Gail index is a discrete random variable with 33 support points and we have 11 5-year age groups for the age-range (30-85). Individuals can give 3 different possible responses to the subjective risk question. We thus have potentially a total of 33*11*3=1,089 moments with which to estimate 8 parameters for each education group. However, some points in the Gail support are not populated in the data and we typically estimate the 8 parameters using between 200 and 400 moments.

As is typical for discrete choice models, the parameters $\left(\frac{V}{\sigma_c}, \frac{c}{\sigma_c}\right)$ are only identified up to scale. We estimate these parameters for each education level separately. Table 9 reports estimates of $\left(\frac{V}{\sigma_c}, \frac{c}{\sigma_c}\right)$ across age.

[Table 9]

The parameters (V, c) are identified using both the overall propensity to screen and how the propensity to screen varies with the subjective risk of cancer. The variation in the subjective risk of cancer is driven by both cancer risk factors and by the age-risk associated with cancer. Thus, the parameter c can be loosely said to fit the propensity to screen overall, while the parameter V can be said to fit the response to variation in risk. Together, they capture how screening demands vary with subjective risk across the population.

The parameters (V, c) are difficult to interpret since they capture variation in the VSL across age, the costs over age as well as the heterogeneity

in costs. All of these terms are difficult to measure in the data and for most we only have very little guidance from theory. We therefore refrain from interpreting these parameters directly. What is important for us here is that they capture the differences in screening propensities conditional on the Gail index and the subjective risk assessments. We consider how well the model fits these patterns next.

5.2.1 Model Fit

Rather than trying to interpret the demand parameters directly, we consider whether the parsimonious structure we have estimated fits observed screening patterns. As described above, we fit about 200-400 moments for each education level using a total of 8 parameters.

Figure 1 shows the predicted and observed fraction screened for all moments for which we have at least 5 observations. As expected, the predicted and the observed fraction screened are placed around the 45 degree line. There is substantial dispersion around the 45 degree line, suggesting that the current model does not capture all the relevant determinants of the screening decision. There is however little evidence in this graph that we systematically over- or underpredict for any particular range of the observed screening shares.

In part, this is certainly driven by the fact that the age-profile is fit quite well and that age has a large influence on screening. Figure 2 shows how the predicted and observed shares screening shares vary with age. These two lines are fairly close. Only at age 75 do we observe a sizeable difference between the predicted and the observed shares screening.

There are however notable differences between predicted and observed screening rates. Table 10 shows how the difference between predicted and observed screening rates by subjective screening risk and education.

[Table 10]

For individuals who judge themselves to be of low subjective breast cancer risk, we observe for all education levels that we underestimate the shares of

individual screening. And, for medium risk we observe that we overpredict the screening rates for all 3 groups. It is not quite clear what generates this pattern.

In the discussion of the multinomial Probit results, we showed that whether a women had cancer in the family played a very large role in determining subjective risk assessments. As shown in columns (3) and (6), controlling for main effects and education interactions of the presence of breast cancer seems to largely eliminate the role of the Gail Index in subjective beliefs. Conditional on family risk, the information contained in the other risk factors and in the functional form of the Gail Index seems to be of little consequence in forming subjective risk judgements. The presence of breast cancer when compared to the age at menarche, parity or the age at first birth is simply much a more salient risk factor.

The salience of family risk has consequence for the fit of the model that is estimated using the full risk model. This "full-risk:" model tends to over underestimate the share of women who are screened conditional on the presence of breast cancer in the family. Figure 3 shows that we substantially underpredict the probability to screen for cancer for individuals with 1 or more direct family members with cancer. For this reason we have also reestimated the information model using only the family risk and reported the estimated σ_s in table 8B. Moving to the "family risk" model leads to a substantial improvement of the estimation procedure in the sense that the model converges for all boot-strapped samples drawn and indeed generally converges more rapidly to the parameter estimates. Qualitatively we find that the signal variance is lower for the family risk model and we also find that the difference in signal noise across education is less pronounced than for the full risk model. Most importantly however, we find again that educated women draw more precise signals compared to less educated women. [11]

[11] We also reestimated the demand parameters using the family risk variables only. These are very similar to the parameters reported for the full risk. We find little value in showing these parameters in a separate table, but they are of course available upon request.

5.3 Results

In the previous Section we presented estimates of the information and the demand parameters of our model. In this Section we use these estimates to generate counterfactual behaviors that allows us to judge whether the differences in the information across education levels are quantitatively important. For both the full-risk and the family risk model, we find find that differences in the information model can almost entirely account for the differences in the responsiveness to risk observed in the data. We also find that the overall gradient in screening across education is mainly due to differences in the demand parameters.

In table 11 we show how responsive individuals of different education levels are to variation in objective medical risk under different counterfactual assumptions about the model parameters. To generate these results we maintained the distribution over the Gail-Index and age within each education group and then generated screening rates using the information and screening model from Section 4 for different values of the parameter vector (σ_s, V, c). Using this data, we then estimate for each education group a Probit relating the screening decisions to a full-set of age-dummies as well as the Gail Index.

[Table 11]

In table 12 we show the analogous results for the family risk model. Here we included dummies for the number of direct relatives with breast cancer. Omitted is the dummy for "no family member" and included is a dummy both for "one family member with breast cancer" and "more than one family member with breast cancer". We again show the coefficient estimates for the effect of these dummies in a probit.

[Table 12]

In both tables, the baseline-estimates (specification (1)) are obtained by using the parameter estimates from the full risk model in Tables 8 and 9

that apply to each education group. The reported coefficients confirm the reduced form result that the less educated respond less to risk factors than the highly educated.

We next investigate whether the differences in the estimated information or demand parameters generate the difference in the response to risk factors across education. For this purpose we we use counterfactual parameter values. In specification (2), we apply the σ_s estimated from the education = 16 to all education classes but maintain the demand parameters as estimated separately for each education class. For both the full risk and the family risk model, we find that this leads the less educated to respond substantially more to risk. We then consider how screening behavior changes if we maintain the estimated information parameters, but endow the less educated with the demand parameters of the most educated. In specification (3) we hold the information quality constant at the base-line values and apply the demand parameters of education levels 16 to all. Clearly, differences in the demand parameters play only a small role in generating differences in the responsiveness to risk factors.[12]

The question then arises whether the observed differences in the quality of information across education classes might be responsible for the overall observed gradient in the screening rates across education categories. Interestingly, this is not the case. Table 13 shows how average screening rates vary across counterfactual states for the full risk model and Table 14 for the family risk model. In column 1 we see the screening rates implies by the estimated parameters. We observe a substantial gradient in screening rates with education. Next we go from this base-line to the counterfactual state when the quality of information among all individuals is equal to the value estimated for those with 16 or more years of education. In this counterfactual state, the difference in screening rates between those with 12 or less years of education and those with 16 or more years of education actu-

[12] Specification (4) than shows the results when both the demand and information parameters estimated using those with more than a college degree are applied to all education classes. By construction the response to risk is similar to the right-most entry in specification (1) for all education levels.

ally widens substantially. The increase in the quality of information for the less educated widens the posterior variance of the expected probability of developing cancer. This has the effect of lowering the expected probability of cancer among some women with low risk factors and these women then refrain from screening. Specification (3) then shows that screening rates overall are highly responsive to changes in the demand parameters. Applying the demand parameters of those with 16 years of education to the entire sample results (while maintaining the information parameters) reverses the screening gradient in education.

[Table 13] [Table 14]

Overall, we find that the differences in quality of information across education generate the observed differences in in how much women respond in their screening decision to the presence of risk factors. The overall screening gradient however seems to be generated by differences in the demand for health across education classes.

6 Discussion

In our analysis we have modeled the effect of education on health information as reducing the noise with which people observe health-relevant information. In our data (and indeed, most health data) however, the information which is known to the econometrician is typically a subset of what the decision maker knows; indeed the data is often self-reported and therefore all individual information observed by the econometrician is also available to the individual. Taken literally then, our model would predict no information-side driven interactions between education and responses to self-reported risk factors. That is, how can smaller responses to the presence of risk-factors which the subject knows about be due to "information side" differences by education?

The answer lies in the interpretation of the information-side effects we estimate; intuitively, we model education not necessarily as improving a person's knowledge of their health characteristics, but improving their un-

derstanding of how these characteristics influence their risk of breast cancer. People may be aware of what risk-factors they have, but unaware of how these risk factors affect the marginal returns to different types of preventative care..That is, a person's noisy signal of their risk-factors in our model is a proxy for differential uncertainty about the risk factors associated with breast cancer, and or what they can do to reduce their risks. This is consistent with Kenkel's finding that more educated individuals were more aware of the specific heath risks associated with smoking, drinking, and less exercise (Kenkel 1991a, 1991b).

Another even greater form of uncertainty may be that people also may have different beliefs about the usefulness of medical science or the relative importance of medical care. Several studies have suggested that differential beliefs and faith in the medical system may explain important health disparities, for example several studies have found lower levels of drug adherence among African Americans controlling for a wide set of demographics (Shenolikar et al. 2006).

Suggestive that both of these types of model-uncertainty may be important in breast-cancer screening, we find two interesting patterns in a recent survey conducted by the San Francisco Mammography Registry (Table 13).

[Table 15]

First, we find that education is highly correlated with a woman's score on a test of accurate knowledge about risk-factors for breast cancer (as identified but those risk-factors included in the Gail model). Second, we find that controlling for other demographic factors, education is highly predictive of a woman's likelihood of agreeing with statements that downplay the relative efficacy of preventative medical care, statement such as: "if a person prays about it, God will protect her from getting cancer", and "if a lump in a woman's breast is not bothersome, there is no need to consult a doctor". This suggests that the noisy-information model we propose is at least partially a proxy for health related model-uncertainty, both on the level of concrete knowledge of risk-factors, and on more meta-cognitive beliefs about

the relative efficacy of medical-science and health care.

7 Conclusion

We have above discussed evidence on allocative efficiency from breast cancer screening. First, we showed that educated individuals are more responsive to the presence of risk factors in both the screening decision and in their judgements about the subjective risk they are facing. These results in themselves are highly suggestive that more educated individuals process health information more efficiently and incorporate this information more readily into their health investment decisions.

We then illustrated the identification problems that make it impossible to separately identify the effect of education due to allocative efficiency from differences in the demand for health by education level using only observational data. To resolve this identification problem, we estimate a structural model of information that allows to generate a joint distribution of subjective risk and objective risk factors from a set of questions about the subjective risk of agents that is contained in the NHIS data. We estimate this information model separately by education level and then use the obtained estimates to estimate the demand parameters governing the screening decision.

Our results are consistent with the reduced from model in that we find that educated individuals are much better informed about the risks they face and are furthermore more responsive to the risks they face when making screening decisions. Our model treats all risk factors symmetrically and while this is an attractive a priori feature of the model, it does seem to fail empirically. The presence of cancer in the family is a risk that seems substantially more salient in the subjective risk assessments of individuals than other risk factors. We do not allow for this feature of the data in our current model and thus underestimate the propensity to screen for individuals with one or more direct relatives with breast cancer. Future work will estimate the same risk model, focusing only on family risk.

Our structural model allows us to examine the counterfactual screening rates obtained if all education groups had the same signal precision as do the

college educated or if they all had the same demand parameters. Our first set of findings suggests that differences in information processing indeed are quantitatively very important for understanding why the more educated are more responsive to the presence of risk factors in screening.

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), "Education and Health: Evaluating Theories and Evidence", forthcoming in J. House, R. Schoeni, G. Kaplan, and H. Pollack, The Effects of Social and Economic Policy on Health, New York: Russell Sage Foundation, 2007.

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.

Gail M. H., et al. (1989) "Projecting individualized probabilites 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.

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.

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

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

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 fo 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.

Shenolikar R.A. et al. (2006) "Race and Medication Adherence in Medicaid Enrollees with Type-2 Diabetes." J Natl Med Assoc. 98;7:1071-7.

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

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

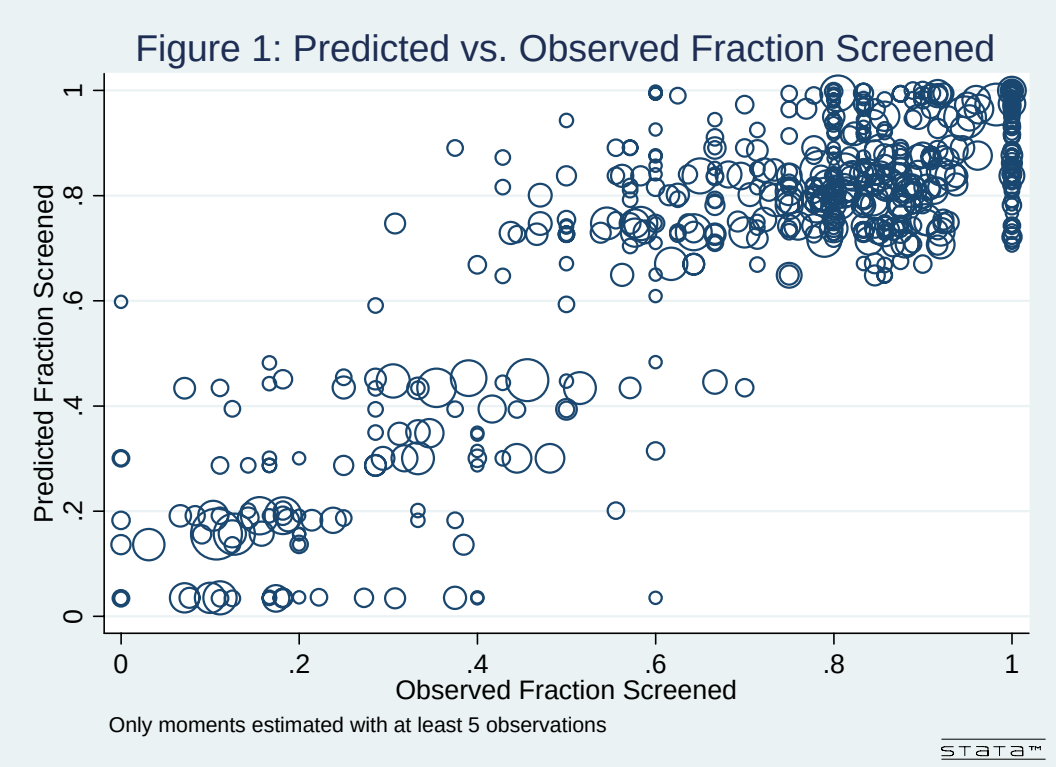
Figure 1: Predicted vs. Observed Fraction Screened
Predicted Fraction Screened
0
.2
.4
.6
.8
1
Observed Fraction Screened
0
.2
.4
.6
.8
1
Only moments estimated with at least 5 observations
Stata™

Figure 2: Predicted and Observed Screening by Age

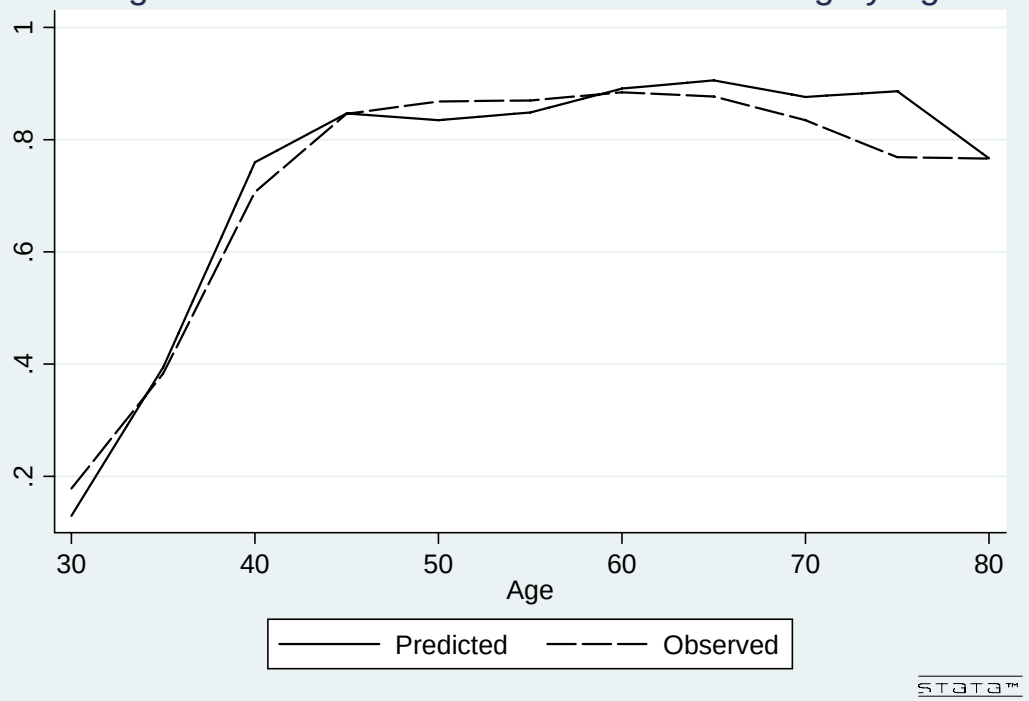

Figure 3: Predicted Screening Not Sensitive Enough to Family Risk

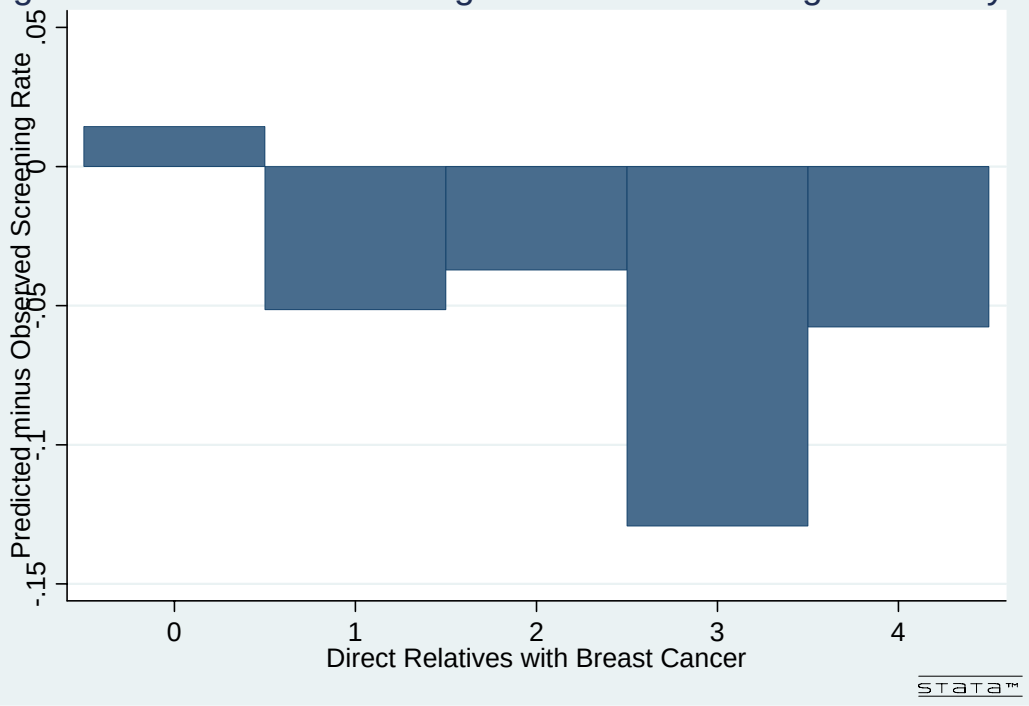

Table 1 Relative risks from the Gail model. Reproduced from Gail et al. (1999)

Risk factor category			Relative risk factor
A. Age at menarche, years			
>13			1
12-13			1.2
<12			1.21
B. Age at first live birth, years	# of first-degree relatives with breast cancer		
<20		0	1
		1	1
		>1	6.8
20-24		0	1.24
		1	2.68
		>1	5.78
25-29 or nulliparous		0	1.55
		1	2.76
		>1	4.91
>29		0	1.93
		1	2.83
		>1	4.17
C. # of breast biopsies			
Age at counseling: <50			
0			1
1			1.27
>1			1.62
Age at counseling, 50+			
0			1
1			1.27
>1			1.62
D. Atypical hyperplasia			
No biopsies			1
At least one biopsy, no atypical hyperplasia			0.93
Atypical hyperplasia in at least one biopsy			1.82

To compute overall relative risk multiply four component relative risk from categories A,B,C,D

Table 2 Summary Statistics

	NHIS 2000		*NHIS 2005*	
Screening Variables				
Ever Had?	0.727		0.755	
# within 6 Years (if >0)	3.81		4.05	
	(2.02)		(2.41)	
Demographic Variables				
White	0.82		0.82	
Black	0.18		0.18	
Age	52.80		53.54	
	(15.68)		(15.28)	
Socio-Economic Variables				
Years of Schooling	13.21		13.50	
	(2.65)		(2.60)	
MSA – Size				
	non-MSA	22.90%	na	
	<250K	9.34%		
	250-500K	12.02%		
	500K-1M	11.83%		
	1-2.5M	23.65%		
	2.5-5M	12.11%		
	>5M	8.14%		
Family Income[1]				
	Aver < 5	2.40	2.40	
		(1.30)	(1.27)	
	% >5	20.67%	21.62%	
	not available	22.20%	20.77%	
Health Care Coverage				
Not Covered	9.98%		8.99%	
Breast Cancer Risk				
Gail Index[2]	1.03		1.09	
	(0.78)		(0.82)	
Parity>0	0.81		0.81	
Age at first Birth (if parity>0)	23.01		23.24	
	(4.99)		(5.14)	
Age at first Menstruation	12.83		12.76	
	(1.77)		(1.70)	
# of direct female relatives with breast cancer				
0	90.11%		88.86%	
1	9.09%		10.16%	
>1	0.80%		0.98%	
Doctor recommended screening[3]	na		53.41%	
Subjective Risk Assessment[4]				
	Low	52.18%	Less likely	34.57%
	Medium	29.49%	About as likely	48.09%
	High	11.47%	More likely	11.64%
	na	6.86%	na	5.71%
Observations	10,379		9,668	

1 Family Income is reported relative to poverty line with 13 categories between 0 and 5. For the summary statistics I assign the mid point to each interval. There is no separate distinction for family incomes above 5 times the poverty line. The percentages in this category as well as those with invalid responses are reported. The analysis uses the income variable as a categorical variable throughout, including invalid responses as a separate category.

2 The Gail Index is a constructed variable using the age at menstruation, age, family cancer history variables, parity, and the age at first birth.

3 (within last 12 months). In 2000 this question was only asked of women who were never screened.

4 In 2000 the subjective risk assessment variable refers to asked whether general subjective risk of cancer was low, medium, or high. In 2005 the question referred specifically to Breast Cancer Specific Risk and asked about likelihood of developing breast cancer relative to average women.

Table 3: Subjective Cancer Risk Assessment and Education

	2000			2005		
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Schooling	-0.045	-0.044	-0.025	-0.055	-0.05	-0.02
	[0.009]**	[0.010]**	[0.009]**	[0.009]**	[0.010]**	[0.010]*
Gail Index	0.017	-0.057	0.004	-0.083	-0.113	0.02
	[0.085]	[0.093]	[0.097]	[0.085]	[0.089]	[0.097]
Years of Schooling x Gail	0.021	0.020	0.001	0.033	0.028	0.000
	[0.006]**	[0.007]**	[0.007]	[0.006]**	[0.007]**	[0.007]
Income x Gail		0.010			0.010	
		[0.007]			[0.006]	
High Income x Gail		0.058			0.102	
		[0.072]			[0.067]	
Breast Cancer in Family			-0.053			-0.523
			[0.240]			[0.253]*
Breast Cancer in Family X Education			0.066			0.117
			[0.018]**			[0.019]**
Observations	9,667	9,667	9,667	9,116	9,116	9,116

Standard errors in brackets. * significant at 5%; ** significant at 1%. Report estimates from ordered probit regressions on individual cancer risk assessment: low, medium, high. In 2000, the question about cancer risk referred to all cancers, whereas in 2005 the question referred to Breast Cancer specifically. All specifications control for ethnicity, age-dummies, ratio of income to poverty (dummies). The Family Risk variable in columns (3) and (6) is a dummy measuring whether a direct relative in the family has ever been diagnosed with breast cancer.

Table 4 Panel 1: Breast Cancer Screening Behavior, Risk, and Education – 2000

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Ever received a Mammogram?					# of Screens in last 6 years				
Years of Schooling	0.020 [0.011]				0.029 [0.012]*	0.046 [0.022]*	0.045 [0.022]*			0.057 [0.023]*
Gail Index	-0.246 [0.108]*	-0.056 [0.141]	-0.345 [0.123]**	-0.389 [0.118]**	-0.236 [0.114]*	-0.510 [0.202]*		-0.050 [0.252]	-0.822 [0.222]**	-0.384 [0.208]
School*Gail Index	0.027 [0.009]**	0.025 [0.009]**	0.016 [0.010]	0.018 [0.009]	0.025 [0.009]**	0.065 [0.015]**	0.061 [0.015]**	0.056 [0.016]**	0.055 [0.017]**	0.053 [0.016]**
Gail Index ^2		-0.028 [0.012]*					-0.060 [0.021]**	-0.059 [0.021]**		
Income*Gail			0.031 [0.009]**	0.034 [0.009]**					0.053 [0.016]**	
(Income>5)*Gail			0.473 [0.111]**	0.509 [0.110]**					0.609 [0.176]**	
Medium subj. Cancer Risk					0.157 [0.036]**					0.302 [0.069]**
High subj. Cancer Risk					0.368 [0.053]**					0.649 [0.097]**
Observations	10,379	10,379	10,379	10,379	9,667	10,234	10,234	10,234	10,234	9,557

Standard errors in brackets.* significant at 5%; ** significant at 1%. Columns 1-5 report estimates of a probit specification with dependent variable: has a woman ever undergone a mammography. Cols 6-10 report estimates of tobit using the left-censored variable: #-mammaograms within last 6 year.
Column 1,6: Baseline with dummies for income, age, ethnicity and race
Column 2,7: + Education dummies, quadratic gail
Column 3,8: + income*gail interaction
Column 4,9: + health insurance
Column 5,10: as in cols 1 and 5 with indicators for self assessed risk.

Table 4 Panel 2: Breast Cancer Screening Behavior, Risk, and Education – 2005

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Ever received a Mammogram?						# of Screens in last 6 years					
Years of Schooling	0.011 [0.013]					0.010 [0.026]	0.018 [0.023]					
Gail Index	-0.139 [0.134]	0.038 [0.175]	-0.203 [0.144]	-0.290 [0.142]*	-0.228 [0.164]	-0.119 [0.146]	-0.463 [0.203]*	0.104 [0.257]	-0.550 [0.218]*	-0.725 [0.217]**	-0.476 [0.204]*	-0.395 [0.214]
School*Gail Index	0.027 [0.010]**	0.022 [0.011]*	0.007 [0.011]	0.010 [0.011]	0.007 [0.013]	0.023 [0.011]*	0.066 [0.015]**	0.054 [0.016]**	0.049 [0.017]*	0.052 [0.017]*	0.044 [0.016]*	0.052 [0.016]**
Gail Index ^2		-0.022 [0.014]						-0.065 [0.019]**				
Income*Gail			0.046 [0.010]**	0.053 [0.010]**	0.046 [0.012]**				0.040 [0.016]*	0.054 [0.016]**	0.029 [0.014]*	
(Income>5)*Gail			0.585 [0.117]**	0.653 [0.118]**	0.536 [0.130]**				0.395 [0.168]*	0.547 [0.167]**	0.354 [0.156]*	
Doctor Recommend					1.109 [0.041]**						2.371 [0.060]**	
Medium subj. Cancer Risk						-0.035 [0.037]						0.146 [0.067]*
High subj. Cancer Risk						0.406 [0.062]**						1.045 [0.104]**
Observations	9668	9667	9667	9667	8859	9116	9568	9568	9568	9568	8777	9041

Standard errors in brackets.* significant at 5%; ** significant at 1%. Columns 1-5 report estimates of a probit specification with dependent variable: has a woman ever undergone a mammography. Cols 6-10 report estimates of tobit using the left-censored variable: #-mammaograms within last 6 year.
Column 1,7: Baseline with dummies for income, age, ethnicity and race
Column 2,8: + Education dummies, quadratic gail
Column 3,9: + income*gail interaction
Column 4,10: + health insurance
Column 5,11: + doc recommendation
Column 6,12: as in cols 1 and 7 with indicators for self assessed risk.

Table 5 Panel 1: Breast Cancer Screening Behavior, Risk, and Education 2000 - Marginal Effects

	(1) Ever?	(2) ever?	(3) ever?	(4) ever?	(5) ever?
Years of Schooling	0.006 [0.003]				
Gail Index	-0.074 [0.032]*	-0.017 [0.042]	-0.102 [0.036]**	-0.115 [0.035]**	-0.082 [0.036]*
School* Gail Index	0.008 [0.003]**	0.007 [0.003]**	0.005 [0.003]	0.005 [0.003]	0.008 [0.003]**
Gail Index ^2		-0.008 [0.004]*			
Income* Gail			0.009 [0.003]**	0.010 [0.003]**	
(Income>5)* Gail			0.141 [0.033]**	0.151 [0.032]**	
Medium subj. Cancer Risk					0.045 [0.010]**
High subj. Cancer Risk					0.098 [0.012]**
Observations	10379	10379	10379	10379	9664

Standard errors in brackets.* significant at 5%; ** significant at 1%. Reported are marginal effects for prbitspecifications in table 2.1 * significant at 5%; ** significant at 1%.
Column 1: Baseline with dummies for income, age, ethnicity and race
Column 2: + Education dummies, quadratic gail
Column 3: + income*gail interaction
Column 4: + health insurance
Column 5: + risk assessments

Table 5 Panel 2: Breast Cancer Screening Behavior, Risk, and Education 2005 - Marginal Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Ever?					
Years of Schooling	0.003					
	[0.003]					
Gail Index	-0.038	0.010	-0.054	-0.076	-0.047	-0.032
	[0.036]	[0.047]	[0.038]	[0.037]*	[0.034]	[0.039]
School*Gail Index	0.007	0.006	0.002	0.003	0.001	0.006
	[0.003]**	[0.003]*	[0.003]	[0.003]	[0.003]	[0.003]*
Gail^2		-0.006				
		[0.004]				
Income*Gail			0.012	0.014	0.010	
			[0.003]**	[0.003]**	[0.002]**	
(Income>5)*Gail			0.156	0.172	0.111	
			[0.031]**	[0.031]**	[0.027]**	
Doctor recommend					0.242	
					[0.009]**	
Avg Subj. Risk						-0.010
						[0.010]
High Subj. Risk						0.094
						[0.012]**
Observations	9668	9667	9667	9667	8859	9116

Standard errors in brackets.* significant at 5%; ** significant at 1%. Reported are marginal effects for probit specifications in table 2.2 * significant at 5%; ** significant at 1%.
Column 1: Baseline with dummies for income, age, ethnicity and race
Column 2: + Education dummies, quadratic gail
Column 3: + income*gail interaction
Column 4: + health insurance
Column 5: + risk assessmen

Table 6: Doctor Screening Recommendations

	Subj Risk		Screening		Doctor Recommendation	
	(1)	(2)	(3)	(4)	(5)	(6)
Gail Index	-0.104	-0.161			0.056	-0.046
	[0.090]	[0.126]			[0.023]*	[0.099]
Education X Gail	0.035	0.04	0.029	0.031		0.008
	[0.007]**	[0.010]**	[0.012]*	[0.015]*		[0.007]
Doctor recommended Mammogram	0.097	0.086	1.108	1.136		
	[0.027]**	[0.252]	[0.041]**	[0.391]**		
Doc Rec X Gail		0.093		-0.095		
		[0.179]		[0.307]		
Doc Rec X Education		0.002		0.015		
		[0.019]		[0.029]		
Doc Rec X Gail X Educ		-0.008		-0.011		
		[0.013]		[0.024]		
Observations	8,372	8,372	8,860	8,860	8,860	8,860

Standard errors in brackets. * significant at 5%; ** significant at 1%. Col 1-2 report ordered probit with dependent variable: Subjective Risk Assessment. Col 3-4 report probit with dependent variable: Ever Screened? Col 5-6 report probit with dependent variable: Did doctor recommend mammogram within last 12 `months?` All specifications control for ethnicity, age-dummies, education dummies, ratio of income to poverty (dummies).

Table 7: Education and Response to Presence of Cancer in Family

	(1) Ever Mammogram	(2) Ever Colonos-copy	(3) Ever Home Blood Stool	(4) Ever Pap Smear	(5) Subjective Breast Cancer Risk	(6) Subjective Colon Cancer Risk
Breast Cancer * Education	0.036 (0.016)**	-0.015 (0.013)	-0.007 (0.013)	0.007 (0.027)	0.100 (0.015)***	-0.015 (0.015)
Colon Cancer * Education	0.028 (0.018)	0.027* (0.014)	-0.007 (0.014)	0.012 (0.029)	-0.003 (0.017)	0.077 (0.017)***
Cervical Cancer * Education	-0.017 (0.031)	-0.004 (0.027)	0.006 (0.027)	0.076 (0.063)	-0.039 (0.026)	-0.033 (0.027)
Observations	20,046	15,315	15,243	18,879	9,116	9,032

Standard errors in parentheses. * significant at 10% ** significant at 5%, *** significant at 1%. Column (1)-(4) report results from probits with dependent Variable indicator for whether an individual has ever received the respective Screen. Colonoscopies and Home Blood Stool tests are screens for colon cancer. Pap Smears are tests for cervical cancer. The results in columns (1)-(4) are from pooled Specifications across both 2000 and 2005. Columns (5) and (6) report results from ordered probits for subjective assessment of "low", "average" or "high" individual cancer risks. These subjective assessments have only been elicited in the 2005 cancer control module for breast and colon cancer, but not for cervical cancer. Specifications include full sets of dummies for age, education, ethnicity, and income fully interacted with year dummies. Furthermore included are indicator variables for whether a direct relative had breast cancer, colon cancer, or cervical cancer. Reported above are the coefficients on interactions of education with these indicator variables.

Table 8A: Dispersion in Individual Signal for Full Risk Model

Completed Years of Schooling	≤12	13-15	≥16
σ_s	3.53 (1.03) [1.30,4.06] {13/60}	1.06 (0.39) [0.69,1.88] {10/60}	0.40 (0.11) [0.27,0.65] {0/60}

Table 8B: Dispersion in Individual Signal for Family Risk Model

Completed Years of Schooling	≤12	13-15	≥16
σ_s	0.49 (0.06) [0.36,0.58] (0/500)	0.32 (0.05) [0.21,0.40] (0/500)	0.27 (0.04) [0.14,0.32] (0/500)

Reported are the point estimates for the standard deviation of the signal by education level. The three measures of dispersion recorded are (from the bottom) the number of boot-strapped samples for which σ_s goes to infinity, the 95% confidence interval and the standard error conditional on observing a finite σ_s.

Table 9: Demand Parameters (Full Risk Model)

Education	≤12		13-15		≥16	
Age	V	c	V	c	V	c
30-34	884	3.68	642	2.24	674	2.45
	(147)	(0.44)	(99)	(0.32)	(104)	(0.23)
35-39	803	4.27	601	2.98	663	3.24
	(127)	(0.58)	(86)	(0.42)	(89)	(0.29)
40-44	729	4.93	571	3.80	658	4.09
	(110)	(0.75)	(80)	(0.58)	(81)	(0.37)
45-49	664	5.68	553	4.70	661	5.03
	(99)	(0.95)	(80)	(0.78)	(79)	(0.45)
50-54	614	6.54	549	5.71	672	6.05
	(92)	(1.17)	(86)	(1.00)	(82)	(0.54)
55-59	583	7.51	560	6.83	693	7.18
	(90)	(1.41)	(95)	(1.24)	(91)	(0.63)
60-64	574	8.60	588	8.07	725	8.43
	(94)	(1.68)	(108)	(1.51)	(102)	(0.74)
65-69	592	9.84	634	9.44	768	9.81
	(103)	(1.97)	(123)	(1.80)	(116)	(0.85)
70-74	642	11.24	700	10.97	823	11.32
	(119)	(2.30)	(141)	(2.12)	(131)	(0.97)
75-79	727	12.79	788	12.66	892	13.00
	(143)	(2.66)	(162)	(2.46)	(150)	(1.11)
80-84	852	14.53	899	14.53	975	14.83
	(175)	(3.05)	(187)	(2.83)	(173)	(1.25)

Reported are the point estimates and bootstrapped standard errors for the demand parameters (V,c) by education and age. These parameters are implied by 4^{th} order polynomials estimated to fit the screening probabilities conditional on education, age, Gail Index, and subjective risk.

Table 10 Predicted minus Observed Screening by Education and Subjective Risk

	Low Risk	Medium Risk	High Risk	Total
≤12	-0.021	0.023	-0.061	-0.004
13-15	-0.035	0.037	-0.006	0.005
≥16	-0.010	0.048	0.026	0.026
Total	-0.023	0.035	-0.019	0.007

Shown are the differences in predicted and observed screening rates conditional on education and subjective risk assessment. The predicted rates are obtained from the model allowing for both family and individual risk. A positive number indicates that the predicted screening rates are smaller than observed screening rates.

Table 11 The response to individual risk: information or demand?

Education	≤12	13-15	≥16
	Baseline Estimates		
(1) With estimated parameters	0.03 (0.003)	0.14 (0.015)	0.35 (0.029)
	Counterfactual Estimates		
(2) With σ_s of educ=16	0.21 (0.018)	0.26 (0.023)	0.35 (0.029)
(3) With (V,s) of educ=16	0.05 (0.005)	0.13 (0.013)	0.35 (0.029)
(4) With (σ_s, (V,s)) of educ=16	0.24 (0.015)	0.26 (0.021)	0.35 (0.029)

Reported are Probit coefficients on the Gail Index. The Probit specification includes a full set of age-dummies. The data for the base-line estimates is generated by taking the (age,Gail) distribution for each education level. We then apply the information and screening model from Section @ as well as the estimated parameters from tables 6 and 7 to generate screening rates conditional on (age,gail,educ,age) and estimate the probit specification from this model. To obtain the counterfactual estimates we took the same (age, Gail) distributions as for the baseline estimates and applied the alternative parameter values from column 1.

Table 12 Accounting for the Education Gradient in Screening: Information or Demand?

Education	≤12	13-15	≥16
	Baseline Estimates		
(1) With estimated parameters	71.9%	77.0%	83.2%
	Counterfactual Estimates		
(2) With σ_s of educ=16	65.9%	73.5%	83.2%
(3) With (V,s) of educ=16	87.1%	84.4%	83.2%
(4) With (σ_s, (V,s)) of educ=16	85.2%	83.4%	83.2%

Reported are mean screening rates by education and across various counterfactual specifications.

Table 13: Education and Beliefs about Breast Cancer

Education	% who agree with		% Scoring above Median on test of Breast Cancer Knowledge
	"Prayer and God can Prevent Cancer"[1]	"Breast Lumps do not need a Doctor"[2]	
High School or Less	31%	4%	28%
More than High School	8%	1%	55%

Data from Sample of women in San Francisco who were screened for Breast Cancer.
1: "If a person prays about it, God will protect her from getting cancer"
2: "If a lump in a woman's breast is not bothersome, there is no need to consult a doctor".