Risk Misperception and Selection in Insurance Markets: An Application to Demand for Cancer Insurance

David S. Hales
University of Nevada, Las Vegas, halesd2@unlv.nevada.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/thesesdissertations
Part of the Economics Commons, Insurance Commons, and the Psychology Commons

Repository Citation
https://digitalscholarship.unlv.edu/thesesdissertations/2359

This Thesis is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
RISK MISPERCEPTION AND SELECTION IN INSURANCE MARKETS:
AN APPLICATION TO DEMAND FOR CANCER INSURANCE

By

David Scott Hales

Science Baccalaureate in Physics
Massachusetts Institute of Technology
1994

Bachelor of Science in Computer Science
National University
1997

Master of Business Administration
University of Hawaii, Manoa
2003

A thesis submitted in partial fulfillment of the requirements for the

Master of Arts – Economics

Department of Economics
Lee Business School
The Graduate College

University of Nevada, Las Vegas
May 2015
We recommend the thesis prepared under our supervision by

**David Scott Hales**

entitled

**Risk Misperception and Selection in Insurance Markets: An Application to Demand for Cancer Insurance**

is approved in partial fulfillment of the requirements for the degree of

**Master of Arts - Economics**

**Department of Economics**

Mary Riddel, Ph.D., Committee Chair
Stephen P. A. Brown, Ph.D., Committee Member
Jeffrey Butler, Ph.D., Committee Member
Hokwon Cho, Ph.D., Graduate College Representative
Kathryn Hausbeck Korgan, Ph.D., Interim Dean of the Graduate College

May 2015
Abstract

Spinnewijn (2013) posits that optimism about risk and the efficacy of risk-reducing effort could cause selection in insurance markets. We test for this using a survey of 474 subjects’ demand for hypothetical cancer insurance. We elicit perceptions of baseline cancer risk and control efficacy and combine these with subject-specific cancer risks predicted by the Harvard Cancer Risk Index to develop measures of baseline and control optimism. We find that only 23 percent of our subjects would purchase a fair insurance contract aligned to their true risk type. Of these subjects, 94 percent also overinvest in prevention, leading to advantageous selection.
Acknowledgements and Declaration of Co-Authored Material

This thesis is based on an article of the same title, co-authored with Professor Mary Riddel and submitted in December 2014 to The Economic Journal for publication consideration. Prior to my involvement in the research project that supports this article, Dr. Riddel envisioned and developed the theoretical framework using the Harvard Cancer Risk Index (HCRI), created a survey design around the HCRI, and wrote the actual Qualtrics online survey that was used in this study. Dr. Riddel also personally conducted the first round of data gathering, soliciting Amazon Mechanical Turk responses from 214 female subjects, in October and November of 2013.

Upon joining the project in October 2013, I performed a literature review and assisted in the data analysis of the initial 214 subjects. I helped extend the project’s scope to include males, adapting the survey to include questions relating to males (in particular, prostate cancer-related questions), in conformance with the relative risk measures contained in the HCRI. I conducted my own survey using Amazon Mechanical Turk in April 2014, on 279 male subjects. In an effort to improve statistics for female subjects, I also conducted a survey of an additional 56 female subjects, using Dr. Riddel’s original female survey, in September 2014.

Both Dr. Riddel and I participated heavily in the analysis, econometric modeling, and hypothesis testing in support of this research. We also each wrote significant portions
of the aforementioned article. In compliance with policies of the Graduate College and by consent of the Department of Economics, I have adapted this article and made additional contributions, to form the body of the thesis.

The portions of the thesis that rely most heavily on Dr. Riddel’s original authorship are Section I, Overview, Section III, Testing for Selection in a Hypothetical Market, and Section IV, Survey Considerations. Section II, Literature Review, is primarily my effort, but with significant contributions and editing by Dr. Riddel. Section V, Indexes for Optimism and Effort, represents one of my primary contributions to this effort. In it I describe my development of a logarithmic, composite set of indexes that we applied in the econometric modeling. Section VI, Selection Model and Results, incorporates significant contributions from both of us. Section VI.C. (Insurance Selection Classification Model) was primarily Dr. Riddel’s effort, and Section VI.D. (Population Classification Predictions and Confidence Intervals) was primarily my effort. In this latter section, Dr. Riddel made the crucial recommendation that I use the Delta Method approach to estimate confidence intervals, and she provided me helpful references in applying this technique. Dr. Riddel and I collaborated and contributed jointly to the results portion of Section VI, as well as to Section VII, Discussion, and to Section VIII, Conclusions.

In all portions of the article and this thesis, including those parts that are my primary authorship, I owe a tremendous debt of gratitude to Dr. Riddel for allowing me to join in
this research effort, and for her guidance, inspiration, mentorship, and on more than one occasion, her patience in correcting and teaching me economic theory and the art of applied econometrics, both as an instructor, as an experienced researcher, and as my thesis advisor.

Additional acknowledgements are due to Dr. Stephen Brown, particularly for his suggestion of adding a final “logic trap” question at the end of the male survey, to detect whether survey respondents were answering questions in a haphazard manner, but also for his enormous and consistent guidance, support, and mentorship since I first took his graduate research seminar, and later as his teaching assistant, where he taught me how to teach. Dr. Hokwon Cho similarly provided helpful guidance in helping me understand bivariate normal distributions, and to apply these in a mathematically correct manner. As my instructor in several statistics courses, he also helped me develop theoretical “bones,” on which best to build strong econometrics “muscle.” Dr. Jeffrey Butler provided helpful behavioral economics insights, both in regards to the current research, and to future possible extensions. Dr. Ian McDonough provided very helpful econometric advice, as well as some very welcome encouragement.

Finally, the comments provided by participants of two conferences I attended with Dr. Riddel, in which she presented earlier versions of the paper, were very helpful in the writing of both the article and this thesis: The Harvard Risk, Perception, and Response Conference, held at the Harvard School of Public Health in March 2014, and the
Behavioral Insurance Workshop, held at the Ludwig-Maximilians University of Munich in December 2014.
To my wife Stephanie, for always being there for me, and for teaching me to celebrate life;

to my daughters Jamie and Julie, for inspiring and encouraging me with their sense of wonder of the world;

and

to my grandson Kaden, for his youthful, innocent spirit that never fails to put a smile on my face.
# Table of Contents

Abstract................................................................................................................................................iii

Acknowledgements and Declaration of Co-Authored Material.........................................................iv

List of Tables.......................................................................................................................................xi

List of Figures......................................................................................................................................xii

I. Overview........................................................................................................................................1

II. Literature Review............................................................................................................................6

III. Testing for Selection in a Hypothetical Market..............................................................................10

IV. Survey Considerations..................................................................................................................13
   IV.A. Risk Perception.......................................................................................................................14
   IV.B. Risk Preference.......................................................................................................................16
   IV.C. Health History and Objective Cancer Risk............................................................................16
   IV.D. Cancer Insurance Demand...................................................................................................17
   IV.E. Cognitive Ability....................................................................................................................18
   IV.F. Demographics.........................................................................................................................18

V. Indexes for Optimism and Effort.....................................................................................................19

VI. Selection Model and Results.........................................................................................................28
   VI.A. Optimism and Preventative Behavior....................................................................................28
   VI.B. Baseline Optimism and Demand for Insurance....................................................................32
   VI.C. Insurance Selection Classification Model...............................................................................35
   VI.D. Population Classification Predictions and Confidence Intervals........................................38
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>OLS Models of Prevention Effort: Dependent Variable is <em>Prevention Effort</em> index</td>
<td>29</td>
</tr>
<tr>
<td>Table 2</td>
<td>Effects of a Change in the Independent Variables in Table 1 on Cancer Risk as a Result of a Change in Prevention Effort</td>
<td>30</td>
</tr>
<tr>
<td>Table 3</td>
<td>Probit Models of Willingness to Pay for Insurance: Dependent Variable Equals One if the Subject Agreed to the Insurance at the Offered Premium</td>
<td>33</td>
</tr>
<tr>
<td>Table 4</td>
<td>Results of the Bivariate Probit Classification Model</td>
<td>37</td>
</tr>
<tr>
<td>Table 5</td>
<td>Subject Population Classified as Adverse or Advantageous Selectors Based on Bivariate Probit Model, with Confidence Intervals derived from the Delta Method and Bootstrap Procedure</td>
<td>46</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1  Histograms of Baseline and Control Optimism for Males and Females...22
Figure 2  Histograms of Estimated Standard Errors of Individual Classification Probabilities.................................................................41
I. Overview

Economists have long posited that asymmetric information with heterogeneous risk types can lead to adverse selection in insurance markets (Rothschild and Stiglitz 1976). Individuals with private information that they are high-risk types tend to buy more coverage than low-risk types. High-risk types will also have higher claims, leading to a positive correlation between insurance coverage and claims. Such positive correlations have been found in some markets, but rejected in others. For example, Puelz and Snow (1994) find evidence of adverse selection in the market for automobile insurance, but Chiappori and Salanie (2000) do not. By contrast, Finkelstein and McGarry (2006) reject the hypothesis of adverse selection in their study of long-term care insurance.

These mixed findings concerning adverse and advantageous selection have led researchers to look for other sources of private information and selection in insurance markets. De Meza and Webb (2001) develop a model where risk aversion leads to advantageous selection as more risk-averse subjects buy more insurance and simultaneously engage in more prevention behavior leading to a negative correlation between coverage and claims. Fang, Keane, and Silverman (2008) examine the market for Medi-gap insurance (a supplement to Medicare). They found that when cognitive ability is controlled for, a negative correlation between coverage and ex-post claims is found, indicative of advantageous selection. They conclude that the correlation arises
because cognitive ability is correlated with both good health and the purchase of health insurance.

A recent paper by Spinnewijn (2013) posits that heterogeneity in risk misperceptions may also affect the relationships between coverage and claims, leading either to advantageous or adverse selection. Spinnewijn (2013) recognizes two dimensions of risk misperception. First, subjects may be relatively “baseline optimistic,” meaning they believe their risk of damages is lower than it actually is. Baseline optimistic subjects are theorized to demand less insurance than their more pessimistic counterparts. “Control optimists” overestimate the risk reductions arising from engaging in preventative activities and avoiding risky activities. As a result, they overinvest in risk-reducing activities relative to their true risk type, leading to lower expected insurance claims.

Assuming a simple model with two insurees with different perceived risk types and incentive-compatible equilibrium contracts, Spinnewijn (2013) shows that if one insuree is more baseline and control optimistic than the other, a positive correlation between coverage and claims will occur. A negative correlation results if the more control pessimistic type is also relatively more baseline optimistic. Thus, depending on the correlation between the control and baseline optimism of the two insurees, either adverse or advantageous selection may result.

Using an online survey of 478 US adults aged 18 and older, we investigate the effects of risk misperception on the willingness to pay for a hypothetical cancer insurance policy
and future expected cancer insurance claims, controlling for risk preferences, cognitive ability, and potentially important demographic variables. We choose cancer insurance because much is known about the role demographics and behavioral choices play in forming cancer risks. We elicit baseline risk perceptions, perceptions of the efficacy of prevention efforts and risk factors related to colon, prostate, and bladder cancer for men and colon, bladder and breast cancer for women. We query subjects about their behaviors that may either reduce or increase risks for one or more of these cancers. We elicit estimates of the subject’s degree of risk aversion using the Holt and Laury’s (2002) multiple price-list elicitation method. We measure cognitive ability using a short intelligence assessment.

To our knowledge, no other research has empirically evaluated the role of risk misperceptions in selection in insurance markets. This is likely because there are few insurable events for which data are available on actual and perceived risk. Thus, the strength of this study is that our measures of cancer-risk misperception rest on applying subjects’ survey responses to the Harvard Cancer Risk Index (HCRI) (Colditz et al. 2000). The HCRI was developed at the Harvard Center for Cancer Prevention, by a working group of “epidemiologists, clinical oncologists, and other Harvard faculty with quantitative expertise focused on cancer and risk assessment” (Ibid.). The HCRI provides quantitative relative risk (RR) factors for each demographic or health behavioral attribute that experts believe bear on the risk of incidence of a given cancer. The HCRI can thus be used to calculate the risk a subject will contract cancer,
conditional on a set of behavioral and demographic traits. By asking subjects to estimate their risk of incidence of each cancer, and comparing that with the HCRI estimates, we derive a measure of their baseline optimism. Similarly, by asking subjects how effective a series of preventative measures are in reducing cancer risk, and how risky a series of unhealthy behaviors are, and then comparing them with the corresponding HCRI RR factors, we derive a measure of their control optimism.

Standard tests for advantageous or adverse selection rest on the sign of the correlation between insurance coverage and claims: positive correlation suggests adverse selection, whereas negative correlation implies advantageous selection. The hypothetical nature of the survey reported in the current paper necessitates a different approach to testing for selection. We first test whether baseline and control optimism influence prevention effort, and thereby affect the risk of contracting cancer. Next, we test whether baseline optimism causes people to under-insure relative to their true risk type. In both models we control for other factors, such as cognitive ability risk aversion, and demographic variables which may influence willingness to pay for insurance and prevention effort. This approach allows us to classify subjects according to whether their willingness to pay for insurance and prevention efforts are high or low relative to their true risk type and health preferences while controlling for other factors that may lead to selection. Given these classifications, we can infer whether positive or negative correlation between coverage and claims will be present in our sample.
Our findings offer strong support for Spinnewijn’s (2013) hypothesis that risk misperception can lead to selection in insurance markets. We find that male subjects who underestimate their likelihood of cancer incidence (baseline optimists) are willing to pay less for full insurance, ceteris paribus. The effect is not present for female subjects. We also show control optimists of both genders engage in more prevention and fewer risky health behaviors, indicating that, on average, their cancer risks and insurance claims will be lower. At the same time, baseline optimism leads subjects to engage in less preventative effort, thereby raising their cancer risk and associated expected claims.

Our classification model indicates that optimism causes over 76 percent of our sample to reject an actuarially fair insurance contract. Of the remaining subjects, 23 percent accept the contract and simultaneously engage in excess prevention; the final 2 percent accept the contract but under-invest in prevention. Thus we show that controlling for risk aversion and cognitive ability, optimism drives most high-risk types out of the market, leading to advantageous selection.
II. Literature Review

Rothschild and Stiglitz (1976) introduced the theory of adverse selection, hypothesizing that even small amounts of asymmetric information in competitive markets can lead to significant distortions of market-clearing prices and quantities. They focused their study on the market for insurance, where they posited that when insurees of heterogeneous risk types have private information about their level of risk, adverse selection may result. Individuals who know they are high-risk types tend to buy more coverage than low-risk types. High risk-types will also have higher claims, leading to a positive correlation between insurance coverage and claims. With the resulting downward-sloping marginal cost curve, the average cost curve is at all times above the marginal cost curve, leading at best to an under-provision of insurance to those with the lowest levels of risk. Depending on the risk premiums individuals place on insurance, a complete unraveling of an insurance offering is possible (see Einav and Finkelstein (2011) for a more detailed discussion).

The seminal paper by Rothschild and Stiglitz (1976) spurred a robust theoretical and empirical literature. Much of the empirical work involved estimating correlations between the amount of insurance coverage and ex-post expenditures on claims predicted by Rothschild and Stiglitz (1976). A positive correlation suggests adverse selection, whereas a negative correlations points to advantageous selection. For example, Chiappori and Salanie (2000) analyzed data from automobile insurance
contracts for young French drivers and found that, when observables are adequately taken into account, no evidence of asymmetric information remains. They concluded that this may be because young drivers do not know their risk types, and older drivers do not know more about their risk types than do the insurance companies. However, Cohen (2005) performed a similar examination of automobile insurance data and found a significant positive correlation between coverage and claims for more experienced drivers, suggesting these more experienced drivers may have learned about their own risk types to a greater extent than had their insurers and less experienced drivers, leading to information asymmetries that result in adverse selection.

De Meza and Webb (2001) noted several previous studies that found either a lack of evidence of adverse selection, or even a negative correlation between coverage and claims. They proposed a model in which an additional factor, risk aversion, plays a key role. They theorized that less risk-averse people are less likely to take precautions, but also less likely to purchase insurance. This then leads to a negative correlation between coverage and risk, and therefore advantageous selection, particularly in the presence of significant administrative costs.

In a similar vein, Fang, Keane, and Silverman (2008) found advantageous selection in the market for Medi-gap insurance (a supplement to Medicare). They found that, controlling only for gender, age, and state of residence (the determinants of policy prices), Medi-gap policyholders spent on average $4,000 per year less on health care
than similar-aged Medicare recipients who do not purchase Medi-gap insurance.

However, when they included a robust set of controls for health, they found that those with Medi-gap spend about $2,000 more than those without Medi-gap. They controlled for additional individual attributes and found that when cognitive ability is controlled for, a negative correlation between coverage and ex-post claims is found, indicative of advantageous selection. They proposed that as cognitive ability is correlated with both good health and the purchase of health insurance, it leads to a negative correlation between Medi-gap coverage and health risk.

Underscoring the sometimes complex dynamics underpinning the demand for insurance, Finkelstein and McGarry (2006) identified multiple forms of private information that can potentially affect the correlation between insurance coverage and risk occurrence. They proposed that it is possible for two or more types of private information to have offsetting effects, leading to behavior that lacks a correlation between risk type and coverage. In the long-term care insurance market, they identified wealth and healthcare preventive activities as being positively correlated with insurance coverage, and negatively correlated with risk.

A range of other studies have produced varying results, which the authors attribute to the particular characteristics of the markets under study. For example, Davidoff and Welke (2004) found evidence of advantageous selection in the reverse mortgage
insurance market. He (2008) found adverse selection in the life insurance market with significant correlation between mortality risk and life insurance coverage.

More recently, Spinnewijn (2013) advanced the idea that risk misperception may also lead to selection in insurance markets. His model assumes two types of risk misperceptions which, acting together, can lead to either adverse or advantageous selection. Policyholders who are “baseline optimistic” believe their risks of experiencing insured events are lower than they actually are; such individuals demand less coverage and engage in less preventative effort, ceteris paribus. Those who are “control optimistic” believe their efforts to mitigate potential negative health effects are more effective than they actually are. Under Spinnewijn’s hypothesis, control optimistic individuals believe that the marginal return to effort is higher than it actually is, and therefore overinvest in effort and hence reduce their expected ex-post claims relative to an individual with accurate or pessimistic views about the return to effort. All other things being equal, an individual who is control optimistic is likely to have lower claims, due to their greater amount of preventative care and avoidance of risky health behaviors. Thus, whether adverse selection is possible rests on the relative influence of baseline and control optimism on insurance demand and prevention effort.
III. Testing for Selection in a Hypothetical Market

To our knowledge, a database that includes subject-level coverage and claims data as well as measures of subjective control and baseline optimism related to cancer risk and insurance does not exist. As such, we take a novel approach to investigating selection in insurance markets. Rather than analyzing historical insurance coverage and claims data, we classify subjects into four classes according to difference in willingness to pay and exertion efforts relative to the corresponding values given their true risk type. We then determine who will purchase an insurance contract with fixed coverage at a given premium and how much effort they will exert. This allows us determine how optimism influences the composition of insured parties and the effort they exert. If the market is dominated by low-risk types who exert high levels of effort, we infer advantageous selection.

Assume a group of risk-averse subjects who are identical in all respects save for their levels of baseline and control optimism. The subjects’ willingness to pay for insurance coverage $R$ given their true risk type is price $P^* = P^f + \theta$ where $P^f$ is the actuarially fair price and $\theta$ is equal to the risk premium the subject is willing to pay. If a subject is baseline pessimistic, then they perceive their risk to be higher than their true risk type, and they will be willing to pay $P^b > P^*$ for coverage $R$. Baseline optimists will be willing to pay $P^f < P^*$. Thus if an insurer offers coverage $R$ at $P^*$, baseline optimists will reject
the coverage thereby underinsuring relative to their true risk type, whereas baseline pessimists will purchase it, considering it to be a bargain.

Of course, whether this leads to adverse or advantageous selection depends on how baseline and control optimism influence behavior. Assume that if the subjects understood the actual efficacy of prevention effort, they would invest $E^*$ in effort. If the combined effects of baseline and control optimism lead the subject to overinvest in effort relative to their true risk type and preferences, then the invest $E^h > E^*$, underinvestment attributable to baseline and control optimism is then $E^l < E^*$.

Define class $C_{ij}$ where $i = 1$ if $E = E^h > E^*$ and 0 otherwise and $j = 1$ if $P = P^k > P^*$ and 0 otherwise. Thus, class $C_{11}$ expends more prevention effort and has a surplus willingness to pay, while $C_{10}$ expends excess effort but underinsures relative to their true risk type. To understand how this classification reveals selection, consider the case where only two classes exist in the market, $C_{10}$ (excess effort, deficient willingness to pay) and $C_{01}$ (deficient effort, excess willingness to pay). As classed, these subjects are equal in all respect save for their level of control and baseline optimism and corresponding effort and willingness to pay. If fair insurance based on the true risk type is offered, $C_{10}$ class will continue to overinvest in effort but will not insure since their willingness to pay is less than the premium offered. They will only enter the market if the price is dropped below $P^*$. The market will be dominated by the $C_{01}$ type since
their willingness to pay exceeds the fair price plus the risk premium. This will cause adverse selection since the type that insures also underinvests in prevention effort.

The outcome will differ depending on the mix of types. Another simple case arises when the market is comprised of solely the $C_{11}$ and $C_{00}$ types. The former type will buy fair insurance and overinvest in effort whereas the latter type will reject the insurance even as they underinvest in prevention effort. Thus the high-risk, low-effort type is driven out of the market by their optimism and the market will be composed of the low-risk type, leading to advantageous selection.
IV. Survey Considerations

We conducted an online survey of 474 men and women aged 18 and over on Amazon Mechanical Turk (AMT). The AMT web service is essentially a labor market designed to match employers who need short tasks completed which require human intelligence to workers willing to complete the task. The tasks, which typically require between 5 and 45 minutes to complete, range from surveys and writing brief product descriptions to transcribing audio recordings. Employers sign up for the service and post task descriptions together with a per-task compensation amount. Employees select tasks using the web as the employer/employee interface.

AMT has become increasingly popular over the past five years with social science and business researchers because of the ease of use of the platform and the streamlined and rapid process for recruiting study volunteers. Buhrmester, Kwang, and Gosling (2011) found that AMT is an inexpensive source for high-quality data. They showed that participants are slightly more diverse than a typical internet sample and much more diverse than a sample based on university students. They also found that the data quality was at least as high as a standard internet or telephone survey design.

¹ Note that we collected survey data from a total of 559 respondents (280 women, and 279 men); of these, we excluded a total of 85 survey responses for several reasons: 1) the survey respondent indicated she or he currently or previously had cancer, rendering the HCRI relative risk factors, and therefore our survey design, unapplicable; 2) the Amazon Mechanical Turk-provided latitude and longitude suggested the respondent was located outside the United States; 3) the respondent’s Amazon Mechanical Turk Identification Number or Internet Protocol address suggested a duplicate response.
For our study, the task was described to potential participants as a survey related to their beliefs about the cancer risks that would take about 20 minutes. Subjects were given between $2.25 and $4.50 to complete the survey. The survey used a split-sample design. One half of the participants began the survey with information about the causes, risk factors, and prevention strategies for one of the three cancers of interest (colon, bladder and breast for women and bladder, colon, and prostate for men).

Following the information section, these subjects began the questionnaire. The other half of the subjects commenced with the questionnaire without any prior information given about cancer risks. We created an indicator variable, info, which we use in our modeling efforts to control for the effects of the information booklet.

The questionnaire has six components, described below:

IV.A. Risk Perception

Risk perception and misperception have two dimensions in the survey. We first queried subjects about their beliefs about the efficacy of cancer-prevention activities and perceptions of the riskiness, in terms of increased cancer risk, of different risky health behaviors. The responses were combined with the expert-assessed efficacy of different activities and used to form measures of control optimism. The second dimension relates to the subject’s view of their own risk of contracting each one of the cancers. Comparing the subjective assessment of risk to the actual risk predicted by the HCRI allows us to calculate a measure of baseline optimism. Below, we briefly describe the
risk-perception elicitation questions. The formulas for the actual calculations for the two types of optimism are described in section V.

Subjects were first asked to grade the decrease (increase) in relative risk in contracting a given cancer, contingent on undertaking specific preventative (risky) activities. The activities considered varied with the cancer. For example, risk factors for bladder cancer included smoking and exposure to chemicals, whereas risks for colon cancer included excessive red meat consumption and a low-calcium diet among others. Prevention activities for colon cancer included regular exercise, taking multivitamins and taking a daily aspirin, among others.

Following the questions about relative risks, subjects were asked to state their personal risk of getting each one of the cancers in their lifetime, compared to the typical subject of their same age and gender. The possible outcomes ranged from zero risk of getting the cancer (Zero. There is no chance of me getting this cancer), to very high risk (very much above average, five times or more above average).

---

2 The relative risks for each preventative activity were presented as both ranges and qualitative descriptors as follows: not effective/does not reduce cancer risk, somewhat effective/reduces risk 10% to 20% below the average person of the same age and gender, moderately effective/reduces risk 30% to 60% below the average, very effective/reduces risk 60% to 80% below the average, and extremely effective/reduces risk by more than 80% below the average.

3 The relative risks for each characteristic or behavior were presented as both ranges of relative risk and a qualitative descriptor as follows no risk increase, small risk increase, risk is higher but less than double the average risk, moderate risk increase to 2 to 4 times the average risk, large risk increase to 4 to 8 times the average risk, and very large risk increase to more than 8 times the average risk.
IV.B. Risk Preference

This section elicited a range for the risk aversion coefficient for the Constant Relative Risk Aversion utility function defined over mortality risks using the sequential multiple price list auction. Details of this aspect of the experiment can be found in Riddel and Kolstoe (2013). Briefly, the subjects read the following text describing the gambles they will face:

_Hypothetical Health Risk:_ Assume you have been diagnosed with a disease that will certainly be fatal in a year without treatment. There are two treatments, but neither is effective 100% of the time. Assume the costs of the treatment are the same, and neither treatment has side effects.

The subjects were then given a sequence of paired lotteries, and asked to select the one they preferred. For example, the first gamble presented was:

_Treatment A means a 30% chance of 8 more years of life and a 70% chance of 2 more years. Treatment B gives a 90% chance of 1 more year (the treatment fails) and a 10% chance of 13.5 more years._

In subsequent gamble pairs, the outcome in treatment B was varied so that $E[A] - E[B]$ gradually decreases, and eventually becomes negative. The analyst notes where the subject switches from preferring lottery A to preferring lottery B, with later switch points indicating higher levels of risk aversion.

IV.C. Health History and Objective Cancer Risk

Subjects were asked a detailed history of their activities, behaviors, and family history for things that may influence their risks of contracting the three cancers of interest. Questions covered their family history of the cancers in question and health related
behaviors such as exercise, vitamin use, smoking, chemical exposure, and alcohol use.

The responses to these questions were used to provide an objective estimate of their risk of getting each of the cancers, using the HCRI from Colditz et al. (2000).  

IV.D. Cancer Insurance Demand

The subjects next faced a single-bounded contingent valuation exercise to determine their demand for cancer insurance. The insurance for males in the sample was described as follows:

Assume that there is an insurance policy available that will cover any and all costs related to the covered cancers. The cancers covered by the insurance are bladder cancer, prostate cancer, and colon cancer. Considering your current budget, would you be willing to pay the following monthly premium for this insurance assuming it covered all related costs including diagnostic testing, office visits for specialists, hospital stays, treatment costs including chemotherapy and radiation, as well as FDA approved experimental treatments. There are no copays or deductibles and you would be able to choose your own doctors and hospitals. Please assume that your current insurance will not cover these cancers and that you will have to pay all of the costs yourself if you get any of these cancers.

Subjects were randomly assigned a bid amount ranging from $5 to $135 per month and asked if they would be willing to pay that amount for the insurance as described.

---

4 Note that the resulting (relative) risk estimates are normed against the U.S. population of persons of the same age and gender. Although we know of no method of estimating cancer risk that can claim to be without error or possible bias, we assume that risk estimates derived using the HCRI methodology are sufficiently accurate to use in estimating subjects’ levels of baseline and control optimism.

5 Women were asked about insurance that covers bladder, colon and breast cancer.
IV.E. Cognitive Ability.

Subjects were asked to answer a series of 7 questions used in the Wonderlic cognitive ability test. The subject scored a one on each question if they gave the correct answer and a zero otherwise. The variable Cognitive Ability was calculated as the sum of the individual scores.

IV.F. Demographics.

Subjects were asked their gender, age, income, ethnicity, education level, and marital status. Variable Age is measured in years, variable Income is measured in thousands of dollars of annual income, while Boolean indicator variables are assigned based on whether one has completed at least a bachelor’s degree (College,) whether one is Married, is Male, is African American (Black), or Asian American (Asian).
V. Indexes for Optimism and Effort

Given subjects’ perceptions of their risk of contracting cancer and their beliefs about the efficacy of prevention efforts, we need to form measures of baseline and control optimism. While there is no generally agreed on formula for combining perceived and actual risks, we believe that any measure allows us to easily understand the degree of optimism in terms of a relative risk i.e. subject believes their risk is half that of their true risk type. Consistent with this thinking, we developed the measures described below.

V.A. Measure of Baseline Optimism

A subject is baseline optimistic if they underestimate their true risk of cancer relative to those in the US population of their own age (and, in the case of breast and prostate cancers, of their same age and gender). Thus, we measure Baseline Optimism by comparing each subject’s stated population-relative risk estimate of incidence for each of the three cancers, with the subject’s “actual” population-relative risk factor (“ARR”).

We calculate the ARR by applying each subject’s responses to demographic, family history, and lifestyle questions in our survey to the risk estimates tabulated for those behaviors in the HCRI. Given each subject’s survey answers, we then estimate the ARR of subject i’s risk of incidence of cancer j as follows:

$$ARR_{ij} = \frac{1}{PD(\hat{\theta}_i)} \prod_{k=1}^{k,(j)} RR_{jk}(\hat{\theta}_i)$$  \hspace{1cm} (1)
where \( k_{ij}(j) \) is the number of relative risk factors for cancer \( j \) identified in the HCRI, \( \delta_j \) is a vector of subject \( i \)'s demographic characteristics, family history, and lifestyle choices, \( RR_{ik}(\delta_i) \) is the HCRI relative risk measure for subject \( i \) for factor \( k \) of cancer \( j \), and \( PD(\delta_i) \) is a population denominator derived from the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Program. As cited in Colditz et al. 2000, the resulting population-relative risk factor gives expert opinion-derived estimates of given subject’s risk of incidence of cancer \( j \), relative to the US population of persons the same age (and for breast and prostate cancers, gender). Thus, a \( ARR_{ij} \) value of 1.0 implies subject \( i \) has an average risk of cancer \( j \) incidence equal to the average of persons in the U.S. of the same age and gender; a value of 2.0 suggests cancer risk that is twice the average, and a value of 0.5 suggests cancer risk that is half the U.S. average.

The survey asked subjects to estimate their risk of contracting each of three cancers (colon, bladder, and breast for women; colon, bladder, and prostate for men), again relative to persons their same age and gender. Consistent with the methodology suggested in Colditz et al. (2000), we structured survey questions to range from “Very much below average risk,” corresponding to a relative risk value of 0.2, to “Very much above average risk,” corresponding to a relative risk value of 5.0. We label subject \( i \)’s

---

6 Note that as we did not have access to the SEER population denominator for prostate cancer, we used an estimate of 1.107372, based on the average (non-normalized) relative risk factors of our sample of 218 men.

7 In addition to the seven levels of relative risk suggested in Colditz et al. 2000, we also allowed survey respondents to select “No risk,” which we code as a relative risk factor equal to 0.01.
stated estimates of relative risk of cancer $j$ as $SRR_{ij}$, and then using values for $ARR_{ij}$ and $SRR_{ij}$, we can create a measure of subjects’ baseline optimism as follows:

$$Baseline\ Optimism_{ij} = \log_2 \left( \frac{ARR_{ij}}{SRR_{ij}} \right).$$

(2)

Here, a value for $Baseline\ Optimism$ of 0.0 implies that a subject’s own estimates of cancer incidence risk (for cancer $j$) are identical to the expert-derived HCRI estimates, based on her responses to survey questions regarding demographic, family, and lifestyle characteristics. A $Baseline\ Optimism$ value of 1.0 indicates that the subject’s estimates of cancer incidence risk are half of the expert-derived value (making her risk perceptions relatively optimistic), and a $Baseline\ Optimism$ value of -1.0 indicates the subject’s risk estimates are twice that of the expert value (making her risk perception relatively pessimistic). Each increase (decrease) of one point in our measure thus has the effect of doubling the amount by which expert risk assessments exceed (are exceeded by) subjects’ own-risk estimates.

Next, we calculate on overall all estimate of each subject’s tendency to exhibit baseline optimism by taking the average of the separate measures for each of the three cancers considered in our study:

$$Baseline\ Optimism_i = \left( \frac{1}{3} \right) \sum_{j=1}^{3} Baseline\ Optimism_{ij} = \left( \frac{1}{3} \right) \sum_{j=1}^{3} \log_2 \left( \frac{ARR_{ij}}{SRR_{ij}} \right).$$

(3)
Figure 1 gives the distribution of *Baseline Optimism* for males and females. Roughly ¾ of each gender in our sample are baseline optimistic. The distribution for females is somewhat higher variance (std. dev.=1.55) than that of males (std. dev.=1.24).
V.B. Measures of Control Optimism

We label a subject as “Prevention Control Optimistic” if she believes that engaging in beneficial activities is more effective in reducing cancer risks than it actually is.

Similarly, we label a subject as “Risk Control Optimistic” if he believes that engaging in a particular risky activity is more likely to lead to cancer than it actually is; we therefore infer that he overestimates his ability to reduce cancer risks by avoiding or curtailing the risky activity in question.

The survey contained a set of questions for each cancer about perceptions of the relative riskiness of different activities that increase or decrease cancer risk. For a given cancer $j$ and beneficial activity $k$, subjects were asked to estimate risk-reducing factors between no risk reduction effect (RR=1.0) and a risk reduction of ten-fold (RR=0.1).

Comparing these estimates with “actual” expert estimates for each cancer and preventative measure associated with each cancer, subject $i$’s level of prevention control optimism is then estimated as:

$$ Prevention\_Control\_Optimism_i = \frac{1}{\sum_{j=1}^{k_{prev}(j)} \sum_{k=1}^{k_{prev}(j)}} \sum_{j=1}^{k_{prev}(j)} \log_2 \left( \frac{APRR_{jk}}{SPRR_{jk}} \right) $$

(4)

where $k_{prev}(j)$ is the number of preventative measures identified in the HCRI for cancer $j$, $APRR_{jk}$ is the HCRI-assessed “actual” post-preventative behavior $k$ relative risk of cancer $j$, and $SPRR_{jk}$ is the subject’s estimates of relative risk of incidence of cancer $j$,
assuming behavior $k$ (with possible responses coded with RR values ranging from 0.1 to 1.0).

By taking the base-2 logarithm of this ratio, and averaging over the total number of preventative measures identified for each of the three cancers in question, we arrive at a measure of optimism exhibited by subject $i$ for a typical preventative measure. A \textit{Prevention_{Control}_{Optimism}} value of 0.0 suggests the subject’s estimates of prevention effectiveness are, on average, equal to the “actual” expert estimates. A measure of 1.0 implies that on average, the subject believes preventative measures are twice as effective as they actually are; a measure of -1.0 implies that on average, the subject believes preventative measures are half as effective as they actually are.

Similarly, but with one crucial difference, we estimate each subject’s level of risk control optimism as follows:

\[
Risk_{Control_{Optimism}} = \frac{1}{\sum_{j=1}^{r_{risk}(j)} \sum_{k=1}^{k_{risk}(j)}} \sum_{k=1}^{k_{risk}(j)} \log_2 \left( \frac{SRRR_{jk}}{ARRR_{jk}} \right) \]

(5)

where $k_{risk}(j)$ is the number of risky activities identified in the HCRI for cancer $j$, $ARRR_{jk}$ is the HCRI-assessed “actual” post-preventative behavior $k$ relative risk of cancer $j$, and $SRRR_{jk}$ is the subject’s estimates of relative risk of incidence of cancer $j$, assuming behavior $k$ (with responses coded with RR values ranging from 1.0 to 5.0). Note that to produce a consistent meaning the ratio between stated and “actual” risk factors is
inverted relative to preventative activities. That is, if a subject’s stated risk estimate for a given cancer and risky behavior is double the “actual” expert relative risk value, she is a Risk Control Optimist for that particular activity and cancer combination.

Finally, we average the values of the two variables for each subject, to arrive at a characteristic level of control optimism for each subject:

\[ Control\ Optimism_i = \frac{1}{2} \left( Prevention\_Control\_Optimism_i + Risk\_Control\_Optimism_i \right) \]  

(6)

An overall Control Optimism measure of 0.0 indicates that the subject accurately assesses the efficacy of prevention efforts. When Control Optimism = 1, the subject believes engaging in preventative measures (avoiding risky activities) is twice as effective in reducing cancer risk than is actually the case; a measure of -1 implies that, on average, the subject believes exerting such effort is half as effective in reducing cancer risk as it actually is.

The distribution of Control Optimism for males and females is given on the right-hand side of Figure 1. Roughly 93% of males and 90% of females are optimistic about prevention activities. Both the male and female Control Optimism distributions have significant right skew, with the male distribution being markedly platykurtic, and the female distribution somewhat less so.

\[ \text{Note that we elected to weight values for prevention and risk control optimism equally in this estimate, rather than weighting by the number of preventative or risk-related attributes for each cancer. We did this to avoid overweighting the influence of risk-related attributes, of which more were identified in the HCRI (29 for women, 24 for men) than were preventative-related attributes (18 for women and 13 for men).} \]

---

8 Note that we elected to weight values for prevention and risk control optimism equally in this estimate, rather than weighting by the number of preventative or risk-related attributes for each cancer. We did this to avoid overweighting the influence of risk-related attributes, of which more were identified in the HCRI (29 for women, 24 for men) than were preventative-related attributes (18 for women and 13 for men).
V.C. Measure of Preventative Effort and Associated Change in Cancer Risk

To estimate the level of effort each subject exerts in relation to cancer-avoiding or cancer-inducing activities, we employ a composite index based on the relative risks from the HCRI for three activities associated with reducing at least one of the three cancers (exercising at least three hours a week, taking a daily vitamin D supplement, and taking a daily baby aspirin) and three risky behaviors (high red meat consumption, high alcohol consumption, and cigarette consumption):

\[
Preventative\text{ Effort}_i = - \sum_{m=1}^{6} \eta_{im} \cdot \log_2 (RR_m)
\]  

(7)

where \( \eta_{im} \) is a boolean operator, indicating whether subject \( i \) engages in preventative/risky behavior \( m \), and \( RR_m \) is the relative risk associated with activity \( m \).

We construct the index such that engaging in risk-reducing activities will contribute a positive value to our index of preventative effort, while engaging in risk-increasing activities will contribute negative values to the index.

Note that to more accurately identify a potential causative effect between risk perception and risk-related behavior, we restrict the activities in our index to those that can be directly controlled by subjects in the near- to mid-term. Thus, we exclude relative risk measures for, say, a subject’s body mass index, as one’s body weight may

---

\(^9\) Note that for cigarette consumption, the HCRI identifies four different levels of risk: 1) non-smoker; 2) smoking less than 1 pack per day; 3) smoking between 1-2 packs per day; 4) smoking more than 2 packs per day.
not be a direct measure of near- to mid-term choices, but rather may be a result of life-
long eating and exercising habits, as well as genetics.

A Preventative Effort index value of 0.0 implies that a subject engages in none of these
six behaviors with bearing on cancer risk, or alternatively that he engages in a
combination of risky and preventative behaviors in such a way that his risk is the same
as if he engaged in none of them. An index value of 1.0 implies that when confronted
with the decision of engaging or not engaging in each of these six behaviors, the
subject’s choices are such that (in aggregate) his risk of incidence of one or more of the
identified three cancers is half what it would be if he engaged in none of these activities.
An index value of -1.0 implies his risk is twice what it would otherwise be. Importantly,
the prevention index can be used to translate differences in prevention effort into
expected differences in cancer risk, using the formula:

\[
\% \Delta \text{Cancer Risk} = 100[2^{-\Delta \text{Preventative Effort}}, -1].
\]  

(8)
VI. Selection Model and Results

As noted above, rather than inferring selection from correlations between actual coverage and claims, we investigate how baseline and control optimism influence the composition and behavior of consumers in our hypothetical cancer-insurance market. To do so, we must first gauge how optimism influences willingness to pay for coverage and prevention effort. In the models results below, we determine if optimism influences cancer-prevention activity, hence cancer risk and expected claims. In the subsequent sub-section, we estimate models of willingness to pay for insurance as a function of optimism and a set of control variables. For both sets of models, we include a model with all subjects as well as models of the individual genders to account for the fact that males and females were asked about a different set of cancers.

VI.A. Optimism and Preventative Behavior

To test whether optimism affects behavior, we examine the relationship between engaging in either risky health behaviors or prevention activities as a function of baseline and control optimism. Prevention and risk-taking behavior is captured in the index for prevention, Prevention Effort, described above. It is possible that prevention effort and optimism are endogenous. We test for endogeneity using the Durbin-Wu-Hausman test (Wooldridge 2003 pg. 506). There is no evidence of endogenity, so we estimate the model using least-squares regression with standard errors corrected for heteroskedasticity. The regressions also control for other attributes of each subject that
are likely to correlate with prevention effort such as cognitive ability, risk preferences, age, income, education, marital status and ethnicity. The results are reported in Table 1:

Table 1. OLS Models of Prevention Effort: Dependent Variable is Prevention Effort index

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Averse</td>
<td>0.002</td>
<td>0.012</td>
<td>0.002</td>
<td>0.018</td>
<td>0.004</td>
<td>0.017</td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>0.006</td>
<td>0.028</td>
<td>-0.020</td>
<td>0.040</td>
<td>0.033</td>
<td>0.040</td>
</tr>
<tr>
<td>Baseline Optimism</td>
<td>-0.112***</td>
<td>0.021</td>
<td>-0.105***</td>
<td>0.032</td>
<td>-0.118***</td>
<td>0.029</td>
</tr>
<tr>
<td>Control Optimism</td>
<td>0.239***</td>
<td>0.072</td>
<td>0.215**</td>
<td>0.095</td>
<td>0.270**</td>
<td>0.105</td>
</tr>
<tr>
<td>Black</td>
<td>-0.116</td>
<td>0.154</td>
<td>-0.109</td>
<td>0.186</td>
<td>-0.117</td>
<td>0.206</td>
</tr>
<tr>
<td>Asian</td>
<td>0.044</td>
<td>0.088</td>
<td>0.014</td>
<td>0.142</td>
<td>0.117</td>
<td>0.112</td>
</tr>
<tr>
<td>Age^2</td>
<td>-.058***</td>
<td>0.019</td>
<td>-0.079***</td>
<td>0.023</td>
<td>-0.033</td>
<td>0.029</td>
</tr>
<tr>
<td>Age</td>
<td>6.34E-04***</td>
<td>2.43E-04</td>
<td>8.65E-04***</td>
<td>2.72E-04</td>
<td>3.59E-04</td>
<td>3.68E-04</td>
</tr>
<tr>
<td>Income ($000)</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>College</td>
<td>0.293***</td>
<td>0.063</td>
<td>0.233***</td>
<td>0.087</td>
<td>0.323***</td>
<td>0.094</td>
</tr>
<tr>
<td>Married</td>
<td>-0.018</td>
<td>0.069</td>
<td>0.004</td>
<td>0.109</td>
<td>-0.038</td>
<td>0.092</td>
</tr>
<tr>
<td>Male</td>
<td>-0.035</td>
<td>0.061</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>C</td>
<td>1.273***</td>
<td>0.359</td>
<td>1.773***</td>
<td>0.420</td>
<td>0.644</td>
<td>0.573</td>
</tr>
<tr>
<td>n</td>
<td>474</td>
<td>218</td>
<td>256</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.151</td>
<td>0.159</td>
<td>0.160</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 0.1, 0.05, and 0.01 levels, respectively.

The first column includes all subjects, while the second and thirds columns estimate models for subsamples of men and women, respectively. Because Prevention Effort is constructed according to a \( \log_2 \) scale, direct interpretation of the coefficients is difficult. To aid the reader in understanding the model results, we have calculated the change in
cancer-mortality risk for changes from a specified baseline relative risk for each of the statistically significant independent variables. The results are reported in Table 2.\textsuperscript{10}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Predictor of Preventative Effort (Model) & Quartile Values \tabularnewline & 25\% & 50\% & 75\% \tabularnewline \hline
Baseline Optimism (All) & -2.0\% & 5.6\% & 13.7\% \tabularnewline Baseline Optimism (Male) & -1.3\% & 4.9\% & 11.5\% \tabularnewline Baseline Optimism (Female) & -2.6\% & 6.1\% & 15.6\% \tabularnewline Control Optimism (All) & -4.2\% & -8.9\% & -13.3\% \tabularnewline Control Optimism (Male) & -4.5\% & -8.5\% & -12.4\% \tabularnewline Control Optimism (Female) & -4.1\% & -9.5\% & -14.6\% \tabularnewline Age (All) – Years & 31.8 years & 46.4 years & 59.6 years \tabularnewline Predicted Marginal Effect & -8.0\% & - & -8.2\% \tabularnewline Age (Male) – Years & 31.3 years & 45.6 years & 58.7 years \tabularnewline Predicted Marginal Effect & -11.4\% & - & -9.9\% \tabularnewline College (All) & - & -18.4\% & - \tabularnewline College (Male) & - & -14.9\% & - \tabularnewline College (Female) & - & -20.0\% & - \tabularnewline \hline
\end{tabular}
\caption{Effects of a Change in the Independent Variables in Table 1 on Cancer Risk as a Result of a Change in Prevention Effort}
\end{table}

According to the models, ethnicity, marital status, income, risk aversion, and cognitive ability are not significant predictors of prevention in any of the three models. Age is not significant in the female model, but has a convex relationship with prevention effort in the all-subjects and male models. The minimum effort level occurs at 45 years’ age in both models, suggesting that as subject’s age, their prevention effort declines until

\footnote{To calculate the percent change in cancer risk attributable to the relevant independent variable, we first calculate the change in the prevention for a change in variable $X_j$ as

$$\Delta \text{Preventative\_Effort}(X_j) = \frac{\partial \text{Effort}}{\partial X_j} \Delta X_j.$$ The change in cancer risk is then

$$\% \Delta \text{Cancer Risk}_{X_j} = 100 \left( 2^{-\Delta \text{Preventative\_Effort}(X_j)} - 1 \right).$$}
about the population median age (about 46 years old in the U.S.), then increases thereafter. Thus, if the average 45-year old male were to change his behavior so as to engage in preventative effort at the same level as the average 31-year old (the 25th percentile of the population age distribution), his age-adjusted risk of cancer would drop by 8%. Similarly, if he were to boost his preventative efforts to match those of the average 58-year old, his age-adjusted cancer risk would fall by 8.2%.

The model results indicate that baseline optimism leads to higher cancer risks for the majority of the subjects in the sample. Subjects with baseline optimism measures in the highest quartile of distribution engage in behaviors that, as a result of this optimism, increase their cancer risk by at least 13.7%. The figure is slightly lower for males (11.5%) and higher for females (15.6%). For the median subject, cancer risks are increased by 4.9% in the all-subjects model. The lowest quartile of subjects, who are baseline pessimistic, actually experience a modest decline in their cancer risk.

We hypothesized that control optimists, believing that preventative activities are more effective than they actually are, would engage in more effort and thereby lower their cancer risks. The model results bear this out. In the all subjects model, the most optimistic 25% of the sample experience a 13.3% or greater decline in their cancer risk as a result of relatively high level of prevention effort. As with baseline optimism, the effect is slightly stronger for females. The effect of control optimism on cancer risk is
still quite large at the median and in the lowest quartile, with cancer risks falling by 8.9% and by as much as 4.2%, respectively, in the all-subjects model.

VI.B. Baseline Optimism and Demand for Insurance

The second component of selection concerns the willingness to pay for full insurance. We estimate three probit regressions (all subjects, male and female) with the dependent variable $\text{Yes} = 1$ if the subject agrees to pay the stated premium for insurance and zero otherwise. The regressors include $\text{Premium}$ (the insurance premium offered to the subject), $\text{Cognitive Ability}$, $\text{Risk Aversion}$, and a set of demographic controls. To investigate the effect of baseline optimism on the demand for insurance, the models include $\text{Baseline Optimism}$ and the $\text{Baseline Optimism} \times \text{Info}$ interaction variable. The interaction variable allows us to test whether the information on cancer risks and causes provided immediately prior completing the survey mitigates the distortionary influence that optimism about one’s cancer risk may have on the insurance-purchase decision. Again, we conducted a Durbin-Wu-Hausman test (Wooldridge 2003 pg. 506) and failed to reject the null hypothesis of exogeneity of $\text{Baseline Optimism}$ in the probit model.

The model results appear in Table 3. In the all-subjects model, the coefficient of the premium amount is negative and statistically significant, indicating that as the plan premium increases, people demand less cancer insurance, all else equal. Older subjects and higher-income subjects have a higher willingness to pay for the insurance than their
younger, lower-income counterparts. Subjects who report that they are of African or Asian descent have a higher willingness to pay than those who self-report as Caucasian or Hispanic. Cognitive ability and risk aversion are not significant in the model. The average willingness to pay for insurance is $51.03 per month.

Table 3. Probit Models of Willingness to Pay for Insurance: Dependent Variable Equals One if the Subject Agreed to the Insurance at the Offered Premium

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Premium</td>
<td>-0.012***</td>
<td>0.002</td>
<td>-0.011***</td>
<td>0.002</td>
<td>-0.014***</td>
<td>0.002</td>
</tr>
<tr>
<td>Risk Averse</td>
<td>0.007</td>
<td>0.025</td>
<td>-0.007</td>
<td>0.038</td>
<td>0.013</td>
<td>0.033</td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>-0.007</td>
<td>0.040</td>
<td>-0.017</td>
<td>0.061</td>
<td>0.016</td>
<td>0.052</td>
</tr>
<tr>
<td>Baseline Optimism</td>
<td>-0.055</td>
<td>0.063</td>
<td>-0.246*</td>
<td>0.129</td>
<td>0.009</td>
<td>0.075</td>
</tr>
<tr>
<td>Info*Baseline Optimism</td>
<td>0.081</td>
<td>0.080</td>
<td>0.319**</td>
<td>0.152</td>
<td>-0.023</td>
<td>0.098</td>
</tr>
<tr>
<td>Black</td>
<td>0.426*</td>
<td>0.260</td>
<td>1.510**</td>
<td>0.604</td>
<td>0.150</td>
<td>0.305</td>
</tr>
<tr>
<td>Asian</td>
<td>0.393*</td>
<td>0.228</td>
<td>0.188</td>
<td>0.345</td>
<td>0.529*</td>
<td>0.318</td>
</tr>
<tr>
<td>Age</td>
<td>0.010**</td>
<td>0.005</td>
<td>0.004</td>
<td>0.009</td>
<td>0.013**</td>
<td>0.006</td>
</tr>
<tr>
<td>Income ($000)</td>
<td>0.005***</td>
<td>0.002</td>
<td>0.007**</td>
<td>0.004</td>
<td>0.005*</td>
<td>0.003</td>
</tr>
<tr>
<td>College</td>
<td>0.131</td>
<td>0.128</td>
<td>0.307</td>
<td>0.196</td>
<td>-0.045</td>
<td>0.175</td>
</tr>
<tr>
<td>Married</td>
<td>-0.137</td>
<td>0.137</td>
<td>-0.271</td>
<td>0.231</td>
<td>-0.131</td>
<td>0.179</td>
</tr>
<tr>
<td>Male</td>
<td>-0.127</td>
<td>0.126</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

The model results for males appear in column 2. The average willingness to pay for males is lower than the full sample at $36.34 per month. The coefficient of the baseline optimism variable is negative and statistically significant (p-value=0.06), whereas the coefficient of the Baseline Optimism*Information interaction variable is positive and statistically significant (p-value=0.04). We infer that the higher the subject’s baseline optimism, the less likely the subjects is to agree to purchase the insurance at the stated
premium amount. As a consequence, the willingness to pay for insurance is lower for male subjects who are overly optimistic about their cancer risks. The effect is significantly attenuated for subjects who received the information on risk and prevention strategies prior to filling out the questionnaire, as evidenced by the positive and significant coefficient of the interaction of the optimism and information variables.

Like the model including all of the subjects, black men are willing to pay significantly more for the insurance than Caucasians. Willingness to pay is increasing in income, with an additional $1000 of income increasing willingness to pay by about $0.65 per month. According to the model, willingness to pay for men is independent of age, marital status, and whether or not they have a college degree. Risk aversion and cognitive ability are also not statistically significant.

Column 3 gives the results of the insurance model for females. As expected, the coefficient of the premium amount is negative and significant, indicating that the higher the premium offered to the subject, the more likely they are to refuse the insurance. The average willingness to pay for insurance implied by the model is $63.67 per month. As with the model for males, cognitive ability and risk aversion are not statistically significant. In contrast with the males in the sample, baseline optimism does not appear to influence female willingness to pay for insurance. Rather, demographic variables seem to be most important. While age did not play a role among males, the coefficient of age in the female sample is positive and significant. Accordingly, women are willing
to pay roughly $1 per month more for insurance as they age one year. As in the male sample, willingness to pay for insurance is increasing in income, with an additional $1000 in income increasing willingness to pay by about $0.35 per month. Asian women are willing to pay more for insurance than those of other ethnicities.

VI.C. Insurance Selection Classification Model

We now turn to an effort to classify the male subjects in our sample according to their excess willingness to pay for insurance and excess prevention effort.\textsuperscript{11} We define the two variables that represent excess prevention and excess willingness that results from optimism, holding all other model variables constant:

\begin{align*}
\text{Excess Prevention Effort}_i = \beta_{CO,PE} \cdot \text{Control Optimism}_i + \beta_{BO,PE} \cdot \text{Baseline Optimism}_i, \\
\text{Excess WTP}_i = \beta_{BO,WTP} \cdot \text{Baseline Optimism}_i,
\end{align*}

where $\beta_{CO,PE}$ and $\beta_{BO,PE}$ are the estimated coefficients of \textit{Control Optimism} and \textit{Baseline Optimism}, respectively, in the prevention effort models, and $\beta_{BO,DEM}$ is the estimated coefficient of \textit{Baseline Optimism} in the insurance demand model. Note that these variables measure effort and willingness to pay for insurance relative to what the subject would engage in if he knew his true risk type. Thus positive (negative) values represent excess (deficient) effort and willingness to pay relative to the true risk type.

We allow our measures of excess willingness to pay and effort to vary with $X_i$, a column vector of individual $i$’s characteristics; $\beta_{prev}$ and $\beta_{wtp}$, column vectors of

\textsuperscript{11} Because we did not find evidence that female demand for insurance is correlated with baseline optimism, we conducted this particular exercise for male subjects only.
parameters; and corresponding measurement and/or observation errors \( u_{\text{prev},j} \) and \( u_{\text{wtp},j} \), as follows:

\[
\begin{align*}
\text{Excess Prevention Effort}_i &= X_i \cdot \beta_{\text{prev}} + u_{\text{prev},i} \\
\text{Excess WTP}_i &= X_i \cdot \beta_{\text{wtp}} + u_{\text{wtp},i}
\end{align*}
\]

For example, the joint probability that the subject falls in Class \( C_{11} \) is then:

\[
p_{11,i} \equiv P(\text{Excess Prevention Effort}_i > 0; \text{Excess WTP}_i > 0) = P(X_i \cdot \beta_{\text{prev}} + u_{\text{prev},i} > 0; X_i \cdot \beta_{\text{wtp}} + u_{\text{wtp},i} > 0) = P(X_i \cdot \beta_{\text{prev}} > -u_{\text{prev},i}; X_i \cdot \beta_{\text{wtp}} > -u_{\text{wtp},i})
\]

If we make the simplifying assumption that the measurement/observation errors are independently and identically distributed, with a bivariate normal distribution and correlation \( \rho \), with \( \sigma_{\text{prev}}^2 > 0 \) and \( \sigma_{\text{wtp}}^2 > 0 \), i.e.

\[
\begin{pmatrix}
\frac{u_{\text{prev},i}}{\sigma_{\text{prev}}} \\
\frac{u_{\text{wtp},i}}{\sigma_{\text{wtp}}}
\end{pmatrix}
\sim N\left[
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix}
\right],
\]

then a consistent estimator for the classification probability is:

\[
\hat{p}_{11,i} = \Phi_2\left(X_i \cdot \hat{\beta}_{\text{prev}}; X_i \cdot \hat{\beta}_{\text{wtp}} | \hat{\sigma}_\text{prev}, \hat{\sigma}_\text{wtp}, \hat{\rho}\right),
\]

where parameters \( \hat{\sigma}_p \), \( \hat{\sigma}_d \), and \( \hat{\rho} \), and parameter vectors \( \hat{\beta}_{\text{prev}} \) and \( \hat{\beta}_{\text{wtp}} \) are estimated using bivariate probit. We then apply the resulting parameter estimates and individual characteristic vector \( X_i \) to (12) to estimate the respective probabilities of subject \( i \) falling into each of the four classes, \( p_{11,i}, p_{10,i}, p_{01,i}, \) and \( p_{00,i} \).\(^{12}\)

\(^{12}\)For clarity’s sake, and without loss of generality, we will restrict our discussion throughout this section to estimates for the \( C_{11} \) classification. Calculations for the three classes (\( C_{10}, C_{01}, \) and \( C_{00} \)) involves a straightforward switching of the indices and the corresponding signs of the probit regressands. Alternatively, we could change the direction of one or more of the inequalities in (10), and then adjust the cdf calculation in (12) accordingly. For example,
One must be cautious when interpreting these classifications. They represent subjects’ excess willingness to pay and excess effort relative to their true risk type, after extracting other sources of heterogeneity in these variables. As such, they give us information about how optimism alone may induce selection, controlling for other factors such as cognitive ability and risk aversion that could also potentially lead to selection in an insurance market. Thus, they are not tests for selection in total, but only represent the possible contribution of optimism to selection. Nevertheless, when aggregated over a sizable population they give helpful insights into the type of insurance purchasing and preventative-related behavior in which members of the population are likely to engage. The results of the bivariate probit model are reported in Table 4.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Prevention Effort Indicator</th>
<th></th>
<th></th>
<th></th>
<th>WTP Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Averse</td>
<td>-0.024</td>
<td>0.035</td>
<td>0.020</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>0.001</td>
<td>0.079</td>
<td>0.002</td>
<td>0.086</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>-0.031</td>
<td>0.477</td>
<td>-0.259</td>
<td>0.606</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0.333</td>
<td>0.345</td>
<td>0.465</td>
<td>0.338</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.039</td>
<td>0.054</td>
<td>-0.006</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td>Age^2</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Income ($000)</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>0.146</td>
<td>0.187</td>
<td>0.257</td>
<td>0.201</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.144</td>
<td>0.215</td>
<td>-0.437*</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rho</td>
<td>0.729</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald chi-squared(18)</td>
<td>16.56</td>
<td></td>
<td></td>
<td>(p-value = 0.553)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Results of the Bivariate Probit Classification Model

\[
\hat{\beta}_{it} = P(X, \hat{\beta}_{it} > -u_{it}; X, \hat{\beta}_{it} < -u_{it}) = P(X, \hat{\beta}_{it} > -u_{it}; X, \hat{\beta}_{it} > -u_{it}) - P(X, \hat{\beta}_{it} > -u_{it}; X, \hat{\beta}_{it} > -u_{it}) \\
= \Phi_2(X, \hat{\beta}_{it}; \sigma_{it}) - \Phi_2(X, \hat{\beta}_{it}; \sigma_{it}, \sigma_{it}, \beta) .
\]
VI.D. Population Classification Predictions and Confidence Intervals

One useful application of this model is to estimate the proportion of our sample that falls into each of the four classifications. Given consistent estimates for individual classification probabilities, derived by applying (12) above, a linear combination of these estimates yields a consistent estimator for population classification ratios:

$$\hat{P}_{11} = \frac{1}{n} \sum_{i=1}^{n} E\left( \hat{P}_{11,i} \right) = \frac{1}{n} \sum_{i=1}^{n} \Phi_2 \left( X_i \cdot \hat{\beta}_{\text{prev}} ; X_i \cdot \hat{\beta}_{\text{wtp}} \mid \hat{\sigma}_{\text{prev}}, \hat{\sigma}_{\text{wtp}}, \hat{\rho} \right),$$

(13)

where $P_{11,i}$ is the estimated proportion of the population that fall in the $C_{11}$ classification.

Construction of confidence intervals on the classification estimator is not nearly as straightforward, due to the fact that the estimator employs a non-linear transformation, the bivariate normal cumulative distribution function $\Phi_2 \left( \right)$. Therefore, we cannot employ a simple linear combination of estimated variances $\hat{\sigma}_{\text{prev}}^2$ and $\hat{\sigma}_{\text{wtp}}^2$ to in turn estimate the variance and thus the standard error $\hat{\sigma}_{\hat{P}_{11}}$ of classification probability $\hat{P}_{11}$.

To solve this problem, we employed two methods: the delta method, and bootstrapping.

---

13 As before, for simplicity’s sake we will restrict our discussion to the $C_{11}$ classification. However, calculation of classification predictions for the other three classes involves a straightforward switching of the indices and the corresponding signs of the probit regressands.
VI.D.1. Using the Delta Method to Construct Classification Confidence Intervals

As outlined in Greene (2012, pp. 68-69), Feiveson (1999), and Oehlert (1992), the delta method can be used to estimate the standard errors of a vector of transformed parameters. Here, we follow and adapt the derivation from Greene for a bivariate probit objective function. We first define

\[ f_i \left( \hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wtp}} \mid X_i, \hat{\sigma}_{\text{prev}}, \hat{\sigma}_{\text{wtp}}, \hat{\rho} \right) = \Phi_2 \left( X_i \cdot \hat{\beta}_{\text{prev}}, X_i \cdot \hat{\beta}_{\text{wtp}} \mid \hat{\sigma}_{\text{prev}}, \hat{\sigma}_{\text{wtp}}, \hat{\rho} \right) \]  

(14)

as a function of the least squares estimators of the two latent biprobit indexes. We take vector \( X_i \), containing observations on each of the \( p = 9 \) regressors for subject \( i \), as well as sample population parameters \( \hat{\sigma}_{\text{prev}}, \hat{\sigma}_{\text{wtp}}, \) and \( \hat{\rho} \), as given and determined.

Then, dropping the exogenous terms for clarity’s sake, and assuming that \( f_i \left( \hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wtp}} \right) \) is both continuous and continuously differentiable at true parameter values \( \beta_{\text{prev}} \) and \( \beta_{\text{wtp}} \), we then define

\[
C_i(\hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wtp}}) = \begin{bmatrix}
\frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wtp}})}{\partial \hat{\beta}_{\text{prev}}} \\
\frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wtp}})}{\partial \hat{\beta}_{\text{wtp}}}
\end{bmatrix}^T
\]

(15)

as a \( 1 \times 2p \) row vector of first derivatives, with respect to each of the \( 2p \) parameters.

For the sake of clarity, the derivation of the partial derivatives specified in (15), for a bivariate normal cumulative distribution function, is shown in detail in Appendix II.
By applying the Slutsky theorem (Greene 2012, pg. 1073) to (15), we then have:

$$\hat{\text{plim}} f_i(\hat{\beta}_{\text{prev}}; \hat{\beta}_{\text{wtp}}) = f_i(\beta_{\text{prev}}; \beta_{\text{wtp}})$$  \hspace{1cm} (16a)$$

and

$$\hat{\text{plim}} C_i(\hat{\beta}_{\text{prev}}; \hat{\beta}_{\text{wtp}}) = \left[ \frac{\partial f_i(\beta_{\text{prev}}; \beta_{\text{wtp}}) }{\partial \beta_{\text{prev}}} \right]^T = \Gamma_i.$$  \hspace{1cm} (16b)$$

To apply the delta method, we then expand function $f_i$ using a first-order Taylor series approximation, and have:

$$f_i(\hat{\beta}_{\text{prev}}; \hat{\beta}_{\text{wtp}}) \approx f_i(\beta_{\text{prev}}; \beta_{\text{wtp}}) + \Gamma_i \times \left[ \hat{\beta}_{\text{prev}} - \beta_{\text{prev}} \right] \hspace{1cm} (17)$$

Then, applying Greene’s derivation (Greene 2012, pg. 69) to the bivariate normal cumulative distribution function, the estimator of the asymptotic covariance matrix is then:

$$\hat{\sigma}_\text{hij}^2 = \text{Est. Asy. Var} \left[ f_i(\hat{\beta}_{\text{prev}}; \hat{\beta}_{\text{wtp}}) \right] = C_i \left[ \begin{array}{ccc} \hat{\sigma}_{\text{prev}}^2 (X^T X)^{-1} & : & \hat{\rho}\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wtp}} (X^T X)^{-1} \\ \hat{\rho}\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wtp}} (X^T X)^{-1} & : & \hat{\sigma}_{\text{wtp}}^2 (X^T X)^{-1} \end{array} \right] C_i^T.$$  \hspace{1cm} (18)
Estimates for $\hat{\sigma}_{\hat{p}_{11,i}}^2$ and $\hat{\sigma}_{\hat{p}_{11,i}} = \sqrt{\hat{\sigma}_{\hat{p}_{11,i}}^2}$ are then readily derived from the observed data.

Figure 2 contains histograms depicting the distribution of estimated standard errors, $\hat{\sigma}_{\hat{p}_{XX,i}}$, for each of the four individual classification probability estimators.

![Histograms of Estimated Standard Errors](image)

Figure 2. Histograms of the Estimated Standard Errors of Individual Classification Probabilities

If we assume the independence of each subject’s estimate $\hat{\sigma}_{\hat{p}_{11,i}}^2$, we can show that the classification estimator meets the conditions for the Liapunov Central Limit Theorem
and the Lindeberg-Feller Central Limit Theorem, and therefore conclude that the distribution of $\frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_{p_{i,1}}^2$ is asymptotically normal (Rao 1973, pg. 127, also Amemiya 1985, pg. 92).

Then, we have:

$$\hat{\sigma}_{p_{i,1}}^2 = Var(\hat{p}_{1i}) = Var\left(\frac{1}{n} \sum_{i=1}^{n} \hat{p}_{1i,j}\right) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left[ Var(\hat{p}_{1i}) + \sum_{j \neq i} Cov(\hat{p}_{1i,1}, \hat{p}_{1j,1}) \right]$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\sigma}_{p_{i,1}}^2 + 0) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \hat{\sigma}_{p_{i,1}}^2 = \frac{1}{n-1} \hat{\sigma}_{p_{1i}}^2$$

$$\Rightarrow \hat{\sigma}_{p_{i,1}}^2 = \frac{\hat{\sigma}_{p_{1i}}^2}{n-1}$$

$$\Rightarrow \hat{\sigma}_{p_{i,1}} = \sqrt{\frac{\hat{\sigma}_{p_{1i}}^2}{n-1}}. \quad (19)$$

Applying the normality assumption, the two end points for a $(1 - \alpha)$ percent confidence interval, for the population classification estimate $P_{11}$, can then be constructed as:

$$P_{11,LB} = \hat{P}_{11} - t_{n-1,\alpha/2} \cdot \hat{\sigma}_{p_{1i}}$$

$$P_{11,UB} = \hat{P}_{11} + t_{n-1,\alpha/2} \cdot \hat{\sigma}_{p_{1i}} \quad (20)$$

The resulting confidence interval estimates for $\alpha = 0.05$ are shown in Table 5, alongside estimates derived from the bootstrapping method.
VI.D.2. Using Bootstrapping to Estimate Classification Confidence Intervals

To motivate our use of bootstrapping, we first define $Z_{11}$ to be the total number of members of sample population $n$ who properly fall within classification $C_{11}$.\(^{14}\) Since the individual classification probabilities are assumed to be independent, we can express $Z_{11}$ as the sum of $n$ Bernoulli variables:

$$Z_{11} = \sum_{i=1}^{n} b_i, \quad \text{where} \quad P(b_i = 1) = p_{11,i}, \quad P(b_i = 0) = 1 - p_{11,i}. \quad (21)$$

Note that if the values for $p_{11,i} = p_0, \forall i \in n$, that is, if individual classification probabilities are all equal, and deterministic (i.e., measured without error), then the probability mass function for $Z_{11}$ would follow a binomial distribution, with mean $np_0$ and variance $np_0(1 - p_0)$ (Ross 2010, pg. 54). From there it would be straightforward to calculate the probability mass function of population classification ratio $P_{11} = \frac{1}{n} \cdot Z_{11}$, and because binomial distributions are asymptotically normal as $n \gg \frac{1}{p_0}$, to apply the Central Limit Theorem and estimate confidence intervals for the distribution of $P_{11}$ (Wackerly 2008, pg. 379).

Since neither of these premises is true, we must find another way of estimating the probability mass function of $Z_{11}$, and therefore to derive confidence intervals for $P_{11}$.

In addition to the Delta Method described in section VI.D.1., we can also employ the bootstrap method.

---

\(^{14}\) As before, we follow the discussion for classification $C_{11}$ only, for simplicity’s sake.
Note that two stages of randomization are required to properly employ bootstrapping in this context: a first stage to randomly assign values to measurement/observation errors $u_{\text{prev},i}$ and $u_{\text{wtp},i}$, and thereby to calculate individual classification probabilities $p_{11,i}$; and, a second stage of randomization to “test” whether each individual Bernoulli variable results in a classification or non-classification, as specified in (21). The number of individuals thus classified then represents one bootstrapped measurement of $Z_{11}$, and by repeating this procedure a number of times, we can estimate the distribution function of $Z_{11}$, and thereby of $P_{11} = \frac{1}{n} \cdot Z_{11}$, enabling us to estimate confidence intervals for the proportion of our sample properly classified in $C_{11}$.

In our first step, we randomly assign values to measurement/observation errors $u_{\text{prev},i}$ and $u_{\text{wtp},i}$ for each individual, using our assumption from (11) that these stochastic error terms are independently and identically distributed, with a bivariate normal distribution and with correlation $\rho: \begin{pmatrix} u_{\text{prev},i} \\ u_{\text{wtp},i} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$.

Once we have selected appropriate values for these error terms, we then use them to calculate the resulting (perturbed) classification probability for each individual:

$$\hat{p}_{11,j} = \Phi \left( \hat{\beta}_{\text{prev}} \cdot X_i + u_{\text{prev},j}^*; \hat{\beta}_{\text{wtp}} \cdot X_j + u_{\text{wtp},j}^* \mid \hat{\sigma}_{\text{prev}}, \hat{\sigma}_{\text{wtp}}, \hat{\rho} \right).$$ (22)
In our second stage of randomization, we generate a separate uniform random variable $v_i \sim Uniform(0,1)$ for each individual, then sum the number of individuals whose individual classification probability $\left( \hat{p}_{i,t}^* \right)$ exceeds their corresponding value for $v_i$. By doing so, we satisfy the Bernoulli condition specified in (21).

Now, as a function defined on stochastic variables, $Z_{i,t}$ is itself stochastic, with its own (unspecified) distribution, which may not necessarily be normal. To construct an interval with confidence level $(1-\alpha)$, we then empirically calculate endpoints

$$Z_{11,\text{LB}} = \arg \max_k \left( \sum_{q=0}^k \text{Count}(Z_{11,q}) \leq \frac{\alpha}{2} \cdot m \right), \text{ and}$$

$$Z_{11,\text{UB}} = \arg \min_k \left( \sum_{q=k}^m \text{Count}(Z_{11,q}) \leq \frac{\alpha}{2} \cdot m \right).$$

(23)

Finally, we construct end points for each respective population classification confidence interval as:

$$P_{11,\text{LB}} = \frac{1}{m} \cdot Z_{11,\text{LB}}$$

$$P_{11,\text{UB}} = \frac{1}{m} \cdot Z_{11,\text{UB}}$$

(24)

The classifications arising from the model, with $m = 80$, are given in Table 5.
Table 5. Subject Population Classified as Adverse or Advantageous Selectors Based on Bivariate Probit Model, with Confidence Intervals derived from the Delta Method and Bootstrap Procedure\textsuperscript{15}

<table>
<thead>
<tr>
<th>Classification</th>
<th>C11: Increased Effort, Increased Ins Demand (Advantageous)</th>
<th>C10: Increased Effort, Decreased Ins Demand (Adverse)</th>
<th>C01: Decreased Effort, Increased Ins Demand (Adverse)</th>
<th>C00: Decreased Effort, Decreased Ins Demand (Advantageous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated Proportion in Class</td>
<td>0.216</td>
<td>0.399</td>
<td>0.013</td>
<td>0.372</td>
</tr>
<tr>
<td>Standard Error on Estimate (Delta Method)</td>
<td>0.0137</td>
<td>0.0081</td>
<td>0.0080</td>
<td>0.0136</td>
</tr>
<tr>
<td>95% Confidence Interval (Delta Method)</td>
<td>0.189 - 0.243</td>
<td>0.383 - 0.415</td>
<td>0.000* - 0.029</td>
<td>0.345 - 0.399</td>
</tr>
<tr>
<td>95% Confidence Interval (Bootstrapping)</td>
<td>0.167 - 0.266</td>
<td>0.330 - 0.467</td>
<td>0.000 - 0.026</td>
<td>0.305 - 0.439</td>
</tr>
</tbody>
</table>

VI.E. Classification Results

Some 77 percent of our male subjects are thus estimated to fall into either class $C_{00}$ or $C_{10}$, indicating that these subjects will reject a policy with premium equal to the fair price plus a risk premium. By contrast, a smaller portion (~22 percent) of our male subjects are estimated to fall into the $C_{11}$ class, implying their pessimism about their cancer risk would lead them to purchase the insurance contract at the fair price plus a premium, while simultaneously engaging in more prevention effort than they would if

\textsuperscript{15} * - Note that for the C01 classification, $\hat{p}_{01} = 0.013$ is of the same order of magnitude as $\hat{\sigma}_{\mu} = 0.008$, and the estimated lower bound for $\hat{p}_{01}$ is actually negative, which has no economic meaning. We are therefore hesitant to conclude that the normality approximation is sufficiently accurate for this particular classification, and take the constructed confidence interval as suggestive only.
they knew their true risk type. Indeed, they will see this policy as a bargain. Only a small number of subjects (~2 percent) are predicted to fall in the $C_{01}$ class, implying purchase of the insurance contract while simultaneously under-investing in preventative measures, leading to adverse selection\textsuperscript{16}.

Thus, controlling for risk aversion and cognitive ability, our sample is consistent with a market that is composed primarily of relatively low-risk types who engage in more prevention than the subjects who refuse the fair insurance. Subjects who are optimistic about their cancer risk, who comprise most of our sample, reject fair insurance. Thus, pessimism about cancer risk and optimism about the efficacy of prevention effort lead to advantageous selection.

\textsuperscript{16} Only 2 percent of our male subjects are estimated to be classified in the C01 quadrant; this is an unsurprising result, considering the strong tendency for all subjects, both male and female, to be both control optimistic and baseline optimistic. In order to be classified in the C01 quadrant, one would have to exhibit both baseline pessimism (leading to higher willingness to pay) and have a control pessimism measure of a relatively large magnitude, enough to overcome the effects of baseline pessimism on preventative effort (in net, leading to decreased exertion). Our results suggest we cannot be confident any of our subjects exhibited such a combination of risk misperceptions.
VII. Discussion

The models show that heterogeneity in risk perceptions can indeed play a key role in selection in insurance markets. For one, we show that 77 percent of our sample are baseline optimistic while 90 percent are control optimistic, indicating that people’s perceptions of risk and prevention efficacy are not well aligned with expert’s assessments. We show that baseline optimism lowers male subjects’ willingness to pay below what it would be if they knew their true risk type. We also show that baseline optimism discourages prevention effort, raising overall cancer risk. Nevertheless, control optimism leads subjects to engage in more effort, thereby lowering their aggregate risk of the cancers in question.

Our classification model indicates that the interaction of baseline and control optimism on subject’s willingness to pay for insurance and prevention efforts results in advantageous selection where high-risk types reject coverage and only low-risk types with high levels of prevention effort are insured. Of course, different samples could give different outcomes. If, for example, our sample was dominated by subjects who were sufficiently pessimistic about prevention efficacy that they under-invested in prevention, they we would have inferred adverse selection rather than the advantageous selection we found here.
One of the perhaps more intriguing findings is the difference between the insurance models for males and females. While baseline and control optimism influence preventative behavior for both genders, the average woman’s willingness to pay for insurance is not a function of baseline optimism. Of course, women and men were queried about a different set of cancers. It could well be that this is the source of this disparity, rather than any inherent differences between how men and women act on their risk perceptions. Others have shown that highly-publicized risks, especially associated with a dreaded disease such as breast cancer, may lead to exaggerated perceptions of disease risk (Slovic 1987). This could be leveraging the results for women.

Another intriguing finding of the insurance model is that the effects of baseline optimism on the demand for insurance were largely nullified for subjects who received the cancer-risk information prior to taking the survey. To further investigate the effect of information, we regressed the optimism variables on information and demographic controls. We found that information did not influence baseline optimism significantly, but information acted to increase the level of control optimism. These are interesting results that we plan to explore further in future papers.

Past research has found evidence that risk aversion and cognitive ability influence demand in some insurance markets [Guiso and Paiella (2005), Barsky et al. (1995), Cohen and Siegelman (2010), Einav and Finkelstein (2011)]. We estimated a simple
insurance model that included the premium and the risk aversion variable as the only covariates, and found that coefficient of the risk aversion variable was positive and statistically significant. Similarly, the coefficient of cognitive ability is positive and statistically significant in an insurance model that excludes all variables except the measure of cognitive ability and the insurance premium. This may be a symptom of multicollinearity among the variables that is elevating the standard errors of the coefficients of risk aversion and cognitive ability when the full range of covariates is included. Thus, we are hesitant to conclude decisively that cognitive ability and risk aversion do not play a role in cancer insurance demand. It could well be that measurement error and/or multicollinearity are leading to an overestimate of their standard errors in the insurance models. Nonetheless, risk aversion and cognitive ability are not significant predictors of prevention effort even in simple models, and therefore the case for any selection arising from these variables in this sample is very weak.
VIII. Conclusions

In this paper, we report the results of a survey of 474 men and women that analyzes willingness to pay for cancer-care insurance, factors that affect the demand for insurance, and variables that influence cancer-prevention activities. In particular, we seek to test whether risk misperception leads to selection, controlling for other possible sources of selection such as cognitive ability and risk aversion.

We offer evidence that supports Spinnewjn ‘s (2013) hypothesis that selection may occur as a result of subjective misperceptions about baseline cancer risks and the efficacy of health-risk reduction activities. Our statistical results indicate that the more optimistic a male subject is concerning his baseline cancer risk, the lower his willingness to pay for cancer insurance. We do not find evidence of this effect with females, however. We also find that subjects (both male and female) who over-estimate the return to preventative behaviors are more likely to invest in preventative effort, thereby lowering their cancer risk and expected associated health-care costs. The pattern of insurance, choice, prevention behavior, and risk misperception can lead to adverse or advantageous selection, depending on the relative influence of control and baseline optimism on behavior.

The models control for other variables, such as risk aversion and cognitive ability that have been shown to lead to selection in insurance markets. We find weak evidence that
these variables influence willingness to pay for insurance, but no evidence that they are
correlated with preventative effort. Still, these relationships may be present but
clouded by multicollinearity and/or measurement error problems. We recommend that
future studies elicit alternative measures of cognitive ability and risk aversion to further
examine their influence on demand for insurance and prevention activities.

This is the first empirical study we know of that investigates risk misperception as a
source of selection. We believe the results reported here are useful to researchers
interested in risk communication, risk perception as well as selection in insurance
markets. That said, there are limitations to the analysis. For one, the data is based on a
hypothetical market so that people never actually purchased or refused the insurance.
People may well make different choices in a hypothetical market, in the context of a
survey, than they make when purchasing actual insurance policies.
Appendix I: Notes on Survey Methodology

As described in the Acknowledgements, we received a total of 549 completed survey responses. An additional 42 respondents began, but did not complete, the surveys. Of the completed surveys we received, we rejected 71 (leaving 478), for the following reasons:

- Duplicate Internet Protocol Address or Amazon Mechanical Turk ID: 28
- Had Cancer: 23
- Latitude/Longitude Outside United States: 11
- Completed Survey in Under 5 Minutes: 9
- Total: 71

As described in the Overview, we conducted the survey using Qualtrics, and solicited respondents through Amazon Mechanical Turk (AMT). Surveys were conducted on separate dates, between October 2013 and September 2014. Separate surveys were issued for colon, bladder, breast, and prostate cancers, and for each of these, a further difference was that half of the surveys contained information booklets at the beginning of the survey, and others did not. Respondents were unaware that the surveys were different; the only difference they were able to perceive was whether a survey was meant for women or men.

Each survey was open for several days, and was cut off automatically by AMT when we reached the target number of desired respondents. We rejected surveys that were completed in under 5 minutes, reasoning that such surveys were likely not carefully read or answered by respondents. We also rejected surveys that originated from the same IP address, or using the same AMT ID, as a previous survey, reasoning that these were likely the same respondent, or possibly a household member. In either case, selection bias could be present if we allowed more than one response per IP address or AMT ID. We rejected surveys from those who have had or currently have cancer, because the HCRI relative risk factors do not apply to such individuals, and hence any measures we could develop for these individuals using the HCRI would be invalid. Finally, we rejected survey responses that originate from outside the United States, as the HCRI relative risk factors are specific to the U.S. population.

A possible source of measurement error that could bias results is that due to respondents who began, but did not complete the survey. As noted above, 42 of the
591 persons who initiated the survey did not complete it. The vast majority of these quit the survey within the first minute, as measured and recorded by Qualtrics.

As cited in Gravelle and Lachapelle (2015), procedures for handling missing data, and for imputing nonresponse in surveys, is addressed in Allison (2001), Little and Rubin (2002), and Rubin (1987). However, the majority of our incomplete surveys contained no responses at all; by design within Qualtrics, we were able to force responses to all questions we posed to respondents. We were left with 549 fully complete surveys, and 42 surveys that contained little or no data at all. Therefore, techniques for imputing nonresponse were not applicable to the majority of the 42 incomplete surveys.

We nevertheless should consider the possibility that these aborted surveys may have led to measurement error in one or more of our estimations. We reason that while some of these terminated surveys may have been due to internet connection or other technical difficulties, which would likely have been uncorrelated with the regressors and therefore unlikely to bias results, the majority of these terminated surveys were likely due to respondent fatigue, or laziness. If “laziness” is then correlated with one of our regressors, such as age, gender, income, or education level, then this would tend to bias our results. However, as the majority of terminated surveys occurred before demographic questions were asked, many in the booklet phase before any questions at all had been asked, we have no way of estimating possible correlations for these subjects. With a 7 percent survey abort rate, we do however conclude that any bias in our results is likely to have been small.
Appendix II: Derivation of Partial Derivative of Bivariate Normal Cumulative Distribution Function

In (15) we reference the partial derivatives of the bivariate normal cumulative distribution functions, \( \frac{\partial f_i(\hat{\beta}_{prev}, \hat{\beta}_{wtp})}{\partial \hat{\beta}_{prev}} \) and \( \frac{\partial f_i(\hat{\beta}_{prev}, \hat{\beta}_{wtp})}{\partial \hat{\beta}_{wtp}} \), which we will develop here.

From (14) we have:

\[ f_i(\hat{\beta}_{prev}, \hat{\beta}_{wtp}) = \Phi_2 \left( X_i \cdot \hat{\beta}_{prev}, X_i \cdot \hat{\beta}_{wtp} \mid \hat{\sigma}_{prev}, \hat{\sigma}_{wtp}, \hat{\rho} \right) \]

From Johnson and Wichern (2007, pg. 151), we have the bivariate normal probability density function:

\[
\phi_2(\psi_{prev,i}, \psi_{wtp,i}) = \frac{1}{2\pi \hat{\sigma}_{prev} \hat{\sigma}_{wtp} \sqrt{1 - \hat{\rho}^2}} \exp \left[ -\frac{1}{2} \left( \frac{\psi_{prev,i}^2 - 2\hat{\rho} \psi_{prev,i} \psi_{wtp,i} + \psi_{wtp,i}^2}{1 - \hat{\rho}^2} \right) \right] 
\]

which we will develop here.

\[ \mu_{prev} \equiv E \left( X_i \cdot \hat{\beta}_{prev} \right), \text{ and } \mu_{wtp} \equiv E \left( X_i \cdot \hat{\beta}_{wtp} \right). \]

In the following derivation, we substitute into (II.1) our parameter estimates \( \hat{\sigma}_{prev}, \hat{\sigma}_{wtp}, \hat{\rho}, \hat{\mu}_{prev} = \frac{1}{n} \sum_{i=1}^{n} X_i \cdot \hat{\beta}_{prev}, \text{ and } \hat{\mu}_{wtp} = \frac{1}{n} \sum_{i=1}^{n} X_i \cdot \hat{\beta}_{wtp}. \)

The bivariate normal probability density function is integrable, leading to the following specification for the bivariate normal cumulative distribution function:

\[
\Phi_2(\psi_{prev,i}, \psi_{wtp,i}) = \frac{1}{2\pi \hat{\sigma}_{prev} \hat{\sigma}_{wtp} \sqrt{1 - \hat{\rho}^2}} \int_{-\infty}^{\psi_{prev,i}} \int_{-\infty}^{\psi_{wtp,i}} \exp \left[ -\frac{1}{2} \left( \frac{x^2 - 2\hat{\rho}xy + y^2}{1 - \hat{\rho}^2} \right) \right] \, dx \, dy
\]

\[
\Rightarrow f_i(\hat{\beta}_{prev}, \hat{\beta}_{wtp}) = \frac{1}{2\pi \hat{\sigma}_{prev} \hat{\sigma}_{wtp} \sqrt{1 - \hat{\rho}^2}} \int_{-\infty}^{\psi_{prev,i}} \int_{-\infty}^{\psi_{wtp,i}} \exp \left[ -\frac{1}{2} \left( \frac{x^2 - 2\hat{\rho}xy + y^2}{1 - \hat{\rho}^2} \right) \right] \, dx \, dy \quad (II.2)
\]
As stated previously, we will take $\hat{\beta}_{\text{prev}}$ and $\hat{\beta}_{wtp}$ to be vectors of estimated coefficients of the latent bivariate probit indexes. Without loss of generality, and for clarity’s sake, we arbitrarily chose a coefficient in vector $\hat{\beta}_{\text{prev}}$, say $\hat{\beta}_j$, with values for $j$ between 1 and $p$, and then calculate the partial derivative of $f_i$ with respect to $\hat{\beta}_j$. Note that $\hat{\beta}_j$ is stochastic, and as the bivariate normal distribution is both continuous and continuously derivable across its domain, we have:

$$\frac{\partial f_i (\hat{\beta}_{\text{prev}}; \hat{\beta}_{wtp})}{\partial \beta_j} = \frac{\partial f_i (\hat{\beta}_{\text{prev}}; \hat{\beta}_{wtp})}{\partial \psi_{\text{prev},i}} \cdot \frac{\partial \psi_{\text{prev},i}}{\partial \beta_j} \quad (II.3)$$

The right-hand side of the product in (II.3) is readily derived as

$$\frac{\partial \psi_{\text{prev},i}}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \left( \frac{X_i \cdot \hat{\beta}_{\text{prev}} - \hat{\mu}_{\text{prev}}}{\hat{\sigma}_{\text{prev}}} \right) = \frac{X_{ij}}{\hat{\sigma}_{\text{prev}}}, \quad (II.4)$$

where $X_{ij}$ is the observation for exogenous variable $j$ made on subject $i$.

We can then apply the Fundamental Theorem of Calculus (Golberg and Cho 2010, pg. 11) to the left-hand side of the product in (II.3):

$$\frac{\partial f_i (\hat{\beta}_{\text{prev}}; \hat{\beta}_{wtp})}{\partial \psi_{\text{prev},i}} = \frac{\partial \Phi_2 (\psi_{\text{prev},i}; \psi_{wtp,i})}{\partial \psi_{\text{prev},i}}$$

$$= \frac{1}{2\pi \hat{\sigma}_{\text{prev}} \hat{\sigma}_{wtp} \sqrt{1 - \hat{\rho}^2}} \cdot \frac{\partial}{\partial \psi_{\text{prev},i}} \int_{-\infty}^{\psi_{wtp,i}} \exp \left[ -\frac{1}{2} \left( \frac{x^2 - 2\hat{\rho} x \psi_{\text{prev},i} + \psi_{\text{prev},i}^2}{1 - \hat{\rho}^2} \right) \right] dx 
\int_{-\infty}^{-\infty} \exp \left[ -\frac{1}{2} \left( \frac{x^2 - 2\hat{\rho} x \psi_{\text{prev},i} + \psi_{\text{prev},i}^2}{1 - \hat{\rho}^2} \right) \right] dx \quad (II.5)$$

Then, rationalizing the exponent in (II.5) we have:

$$-\frac{1}{2} \left( \frac{x^2 - 2\hat{\rho} x \psi_{\text{prev},i} + \psi_{\text{prev},i}^2}{1 - \hat{\rho}^2} \right) = -\frac{1}{2} \left( \frac{x^2 - 2\hat{\rho} x \psi_{\text{prev},i} + \left( \hat{\rho}^2 + (1 - \hat{\rho}^2) \right) \psi_{\text{prev},i}^2}{1 - \hat{\rho}^2} \right)$$

56
\[
-\frac{1}{2} \left( x^2 - 2\hat{\rho} \psi_{prev,i} y + \hat{\rho}^2 \psi_{prev,i}^2 \right) + \frac{1}{1-\hat{\rho}^2} \psi_{prev,i}^2 = -\frac{\psi_{prev,i}^2}{2} - \frac{1}{2} \left( \frac{x - \hat{\rho} \psi_{prev,i}}{1-\hat{\rho}^2} \right)^2
\]

Substituting back into (II.5), we have:

\[
\frac{\partial f_i(\hat{\mathbf{b}}_{prev,i}^*; \hat{\mathbf{w}}_{wp})}{\partial \psi_{prev,i}} = -\frac{1}{2} \left( \frac{x - \hat{\rho} \psi_{prev,i}}{\sqrt{1-\hat{\rho}^2}} \right)^2.
\]

Now, changing variables of integration and substituting into (II.6), let

\[ u = \frac{x - \hat{\rho} \psi_{prev,i}}{\sqrt{1-\hat{\rho}^2}} \Rightarrow du = \frac{dx}{\sqrt{1-\hat{\rho}^2}} \Rightarrow dx = \sqrt{1-\hat{\rho}^2} \cdot du \]

\[
\Rightarrow \frac{\partial f_i(\hat{\mathbf{b}}_{prev,i}^*; \hat{\mathbf{w}}_{wp})}{\partial \psi_{prev,i}} \frac{1}{2\pi \sqrt{1-\hat{\rho}^2} \hat{\sigma}_{prev} \hat{\sigma}_{wp}} e^{-\psi_{prev,i}^2/2} \cdot \left[ \psi_{prev,i} \int_{-\infty}^{x} \exp \left( -\frac{1}{2} \left( \frac{x - \hat{\rho} \psi_{prev,i}}{\sqrt{1-\hat{\rho}^2}} \right)^2 \right) \right] \frac{dx}{\sqrt{1-\hat{\rho}^2}}
\]

\[
= \frac{1}{2\pi \sqrt{1-\hat{\rho}^2} \hat{\sigma}_{prev} \hat{\sigma}_{wp}} e^{-\psi_{prev,i}^2/2} \cdot \left[ \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} \cdot du \right] \frac{\phi_i(\psi_{prev,i}) \cdot \Phi_i(A)}{\hat{\sigma}_{wp}}
\]

\[
\Rightarrow \frac{\partial f_i(\hat{\mathbf{b}}_{prev,i}^*; \hat{\mathbf{w}}_{wp})}{\partial \psi_{prev,i}} = \frac{\phi_i(\psi_{prev,i}) \cdot \Phi_i(A)}{\hat{\sigma}_{wp}},
\]

where \[ A = \frac{1}{\sqrt{1-\hat{\rho}^2}} \left( \psi_{wp,i} - \hat{\rho} \psi_{prev,i} \right). \]

Combining results from (II.4) and (II.7) into (II.3), we have:
\[
\frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wp}})}{\partial \beta_j} = \frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\psi}_{\text{wp}})}{\partial \psi_{\text{prev},j}} \cdot \frac{\partial \psi_{\text{prev},j}}{\partial \beta_j} = \left[ \frac{\phi_i(\psi_{\text{prev},j}) \cdot \Phi_1(A)}{\hat{\sigma}_{\text{wp}}} \right] \cdot \left[ \frac{X_i}{\hat{\sigma}_{\text{prev}}} \right]
\]

\[
\Rightarrow \frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\psi}_{\text{wp}})}{\partial \beta_j} = \frac{X_i \cdot \phi_i(\psi_{\text{prev},j}) \cdot \Phi_1 \left( \frac{\psi_{\text{wp},i} - \hat{\rho} \psi_{\text{prev},i}}{\sqrt{1 - \hat{\rho}^2}} \right)}{\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wp}}}.
\]

where again for clarity we have substituted:

\[
\psi_{\text{prev},j} = \frac{X_i \cdot \hat{\beta}_{\text{prev}} - \hat{\mu}_{\text{prev}}}{\hat{\sigma}_{\text{prev}}}, \quad \psi_{\text{wp},i} = \frac{X_i \cdot \hat{\beta}_{\text{wp}} - \hat{\mu}_{\text{wp}}}{\hat{\sigma}_{\text{wp}}}.
\]

Since \( \hat{\beta}_j \) was chosen arbitrarily from \( \hat{\beta}_{\text{prev}} \), we can similarly derive the other partial derivatives in \( \frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wp}})}{\partial \beta_{\text{prev}}} \):

\[
\frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wp}})}{\partial \beta_{\text{prev}}} = \frac{1}{\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wp}}} \left[ X_i \cdot \phi_i(\psi_{\text{prev},j}) \cdot \Phi_1 \left( \frac{\psi_{\text{wp},i} - \hat{\rho} \psi_{\text{prev},i}}{\sqrt{1 - \hat{\rho}^2}} \right) \right]^T
\]

\[
\Rightarrow \frac{\partial f_i(\hat{\beta}_{\text{prev}}, \hat{\beta}_{\text{wp}})}{\partial \beta_{\text{prev}}} = \frac{1}{\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wp}}} \cdot \phi_i(\psi_{\text{prev},j}) \cdot \Phi_1 \left( \frac{\psi_{\text{wp},i} - \hat{\rho} \psi_{\text{prev},i}}{\sqrt{1 - \hat{\rho}^2}} \right) \cdot X_i.
\]

Now, by symmetry we can repeat steps described above in (II.3) through (II.9) for any arbitrary \( \hat{\beta}_k \) in \( \hat{\beta}_{\text{wp}} \), yielding:
\[
\frac{\partial f_i(\hat{\beta}_\text{prev}^{\prime}, \hat{\beta}_\text{wtp}^{\prime})}{\partial \beta_k^r} = \frac{X_{ik}}{\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wtp}}} \cdot \phi_i(\psi_{\text{wtp},i}) \cdot \Phi_1 \left( \frac{\psi_{\text{prev},i} - \hat{\rho} \psi_{\text{wtp},i}}{\sqrt{1 - \hat{\rho}^2}} \right)
\]  

(II.10)

and

\[
\frac{\partial f_i(\hat{\beta}_\text{prev}^{\prime}, \hat{\beta}_\text{wtp}^{\prime})}{\partial \hat{\beta}_\text{wtp}^T} = \frac{1}{\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wtp}}} \cdot \phi_i(\psi_{\text{wtp},i}) \cdot \Phi_1 \left( \frac{\psi_{\text{prev},i} - \hat{\rho} \psi_{\text{wtp},i}}{\sqrt{1 - \hat{\rho}^2}} \right) \cdot X_i
\]  

(II.11)

Finally, substituting (II.9) and (II.11) into (15), we arrive at

\[
C_i(\hat{\beta}_\text{prev}^{\prime}, \hat{\beta}_\text{wtp}^{\prime}) = \left[ \begin{array}{c} \frac{\partial f_i(\hat{\beta}_\text{prev}^{\prime}, \hat{\beta}_\text{wtp}^{\prime})}{\partial \beta_{\text{prev}}}^T \\ \frac{\partial f_i(\hat{\beta}_\text{prev}^{\prime}, \hat{\beta}_\text{wtp}^{\prime})}{\partial \hat{\beta}_\text{wtp}^T} \end{array} \right]^T = \frac{1}{\hat{\sigma}_{\text{prev}} \hat{\sigma}_{\text{wtp}}} \cdot \left[ \begin{array}{c} \phi_i(\psi_{\text{prev},i}) \cdot \Phi_1 \left( \frac{\psi_{\text{prev},i} - \hat{\rho} \psi_{\text{wtp},i}}{\sqrt{1 - \hat{\rho}^2}} \right) \cdot X_i \end{array} \right],
\]  

(II.12)

for which values can readily be calculated given observation vector \( X_i \), and estimates previously derived for \( \hat{\sigma}_{\text{prev}} \), \( \hat{\sigma}_{\text{prev}} \), and \( \hat{\rho} \); the resulting vector \( C_i \) can then be substituted into (18) to yield a consistent estimate for \( \hat{\sigma}_{\text{p11}}^2 \).
Appendix III: Copy of Qualtrics Survey

In the following twenty pages we show a copy of the survey used for men, for colon cancer. Similar surveys were used for women for colon cancer, and for both women and men for the other three cancers, as appropriate. Note that approximately half of survey respondents were presented with information on one of the three cancers, and the other half were not.
### Introduction

We are conducting an academic survey about people’s knowledge of the risks of cancer. The survey is designed to get an understanding of your knowledge of the things that cause cancer and the things that you do to increase or decrease your risk of cancer. We are particularly interested in prostate, colon, and bladder cancer. The survey also asks questions that will help us understand how demographic variables such as age and income affect knowledge. Some of the questions may seem strange or unusual, but they are all necessary for the research.

The survey should take 15-20 minutes. At the end of the survey, you will be given a code to get credit for Amazon Turk. You will not receive the code unless you complete all of the survey. Please take the time to answer each question and do not rush. Also, *this survey is being conducted of Men aged 18 and up, living in the United States of America*. Please do not complete this survey if you do not live in the United States, or are not a man. Also, if you completed a similar version of our survey in the past week, you are not eligible to complete this (slightly different) survey. We will certify the uniqueness of your User ID as well as your geographic location, when certifying you for payment.

All of the information you give will be completely confidential and your name or anything that could personally identify you will never be attached to any of the responses. Please answer as truthfully as you can.

Click the arrow to start the survey.

### Colon Booklet

The survey has two parts. The first section contains several pages that give specific information about colon cancer risks. Please read these pages carefully. Once you have carefully read the information, proceed to the next section where the question component of the survey begins.

### Colon Cancer

#### What is the colon?

The colon is the main part of the large intestine, which is the long, muscular tube that food passes through during digestion.

#### What is colon cancer?

Colon cancer occurs when cells in the colon grow out of control. A group of abnormal cells together can form a growth in the colon called a polyp. If not removed, cells in the polyp can continue to grow, turn into cancer, and spread.

#### How common is colon cancer?

Colon cancer is the fourth most common cancer in the US, and it is the number one cause of cancer death in nonsmokers.

#### Who is at risk of getting colon cancer?

Anyone can get colon cancer; it strikes both men and women. Colon cancer is more common in people over 50, and people with a family history of colon cancer have a higher chance of getting the disease.
Who should get regular screening tests for colon cancer?

Everyone (men and women) should be screened regularly for colon cancer starting at age 50. People who have a family history of colon cancer and people who have certain disease, like inflammatory bowel disease, may need to start screening before age 50. Getting regular screening tests is the single best way to lower colon cancer risk. Screening tests can prevent colon cancer by finding polyps. Polyps are small abnormal growths that can sometimes turn into cancer. If they are found early, polyps can be removed before cancer develops. If cancer has already developed, screening can help find it at an early stage, when it can be most successfully treated.

How do you prevent colon cancer?

Get regular screening tests beginning at age 50 (or earlier if you have a family history.)

Be physically active for at least 30 minutes every day

Maintain a healthy weight

Limit the amount of alcohol you drink

Limit red meat to less than 3 servings a week

Take a multivitamin with folate every day

Make sure you get enough calcium and vitamin D

Take an aspirin every day (check with your doctor first)

Risk Factors for Colon Cancer

Age and colon cancer
The risk of colon cancer goes up with age. Most cases are diagnosed in people over the age of 50. The average age the disease is found is 73.

Height and colon cancer
Tall people have a higher risk of colon cancer. Researchers don’t know exactly why, but it may be related to the fact that tall people grow more. Some of the same hormones and other factors that make people grow may also increase the chance that dividing cells become abnormal and turn cancerous.

Weight and colon cancer
People who maintain a healthy weight have a lower risk of colon cancer. One reason may be that fat tissue affects different hormone levels in the body. Too much fat tissue can lead to higher hormone levels and increase the risk of cancer.
People who maintain a healthy weight also have a lower risk of kidney cancer, heart disease, pancreatic cancer, diabetes and stroke. And women have a lower risk of breast cancer, and uterine cancer.
Physical activity and colon cancer
People who are physically active for at least 30 minutes a day have a lower risk of colon cancer, possibly because physical activity affects hormone levels and other growth factors in the body. Being physically active is one of the best ways to help maintain a healthy weight. Physically active people also have a lower risk of breast cancer, heart disease, diabetes and stroke. Even just 30 minutes of moderate activity (like walking) daily can decrease your risk of disease.

Red meat and colon cancer
People who eat less than 3 servings of red meat per week have a lower risk of colon cancer. Avoiding processed meats, like hot dogs, ham, and salami, may be especially good. One possible explanation is that cooked and processed meat may contain chemicals that can cause cells to become cancerous.

Red meat includes beef, pork, veal and lamb. 1 serving is 4 ounces, about the size of a deck of cards.

Alcohol and colon cancer
People who drink more than 1 drink of alcohol per day have a higher risk of colon cancer. (A drink is a can of beer, a glass of wine, or a shot of hard liquor.) There are many possible reasons for this. For example, alcohol may cause abnormal cell changes in the body that can lead to cancer. Alcohol can also decrease levels of folate (folic acid) in the body. Folate is a B-vitamin that helps keep cells from becoming cancerous.

Men and women who limit alcohol also have a lower risk of high blood pressure and stroke. In addition, women also have a lower risk of breast cancer.

However, drinking a moderate amount (not more than 1 drink/day for women or 2 drinks/day for men) has benefits too. People who drink a moderate amount have a lower risk of heart disease and diabetes.

Multivitamins and colon cancer
People who take a multivitamin with folate (folic acid) every day have a lower risk of colon cancer. Folate is a B-vitamin that can help protect cells from abnormal changes that lead to cancer. Multivitamins with folate are an important sources of this vitamin. Folate is also found in different fruits, vegetables, grains, nuts and legumes.

Folate doesn’t just protect against colon cancer. People who take a multivitamin with folate can have a lower risk of heart disease and stroke. And it lowers the risk of birth defects when taken by women before pregnancy or during the early stages.

More Risk Factors

Calcium and colon cancer.
People who don’t get enough calcium have a higher chance of developing colon cancer. Calcium is a mineral that is important for healthy bones, muscles and nervous system. Calcium can also help protect against abnormal growths in the colon (polyps) and colon cancer. Calcium in the diet comes from different sources, like dairy foods, some leafy green vegetables (such as broccoli and collards), and fortified foods. Vitamin supplements also provide calcium. Research suggests that about 700 mg/day helps lower the risk of colon cancer, and more is not necessarily better. People who don’t eat dairy products and other calcium-rich foods may want to take a calcium supplement.

Vitamin D and colon cancer
People who have higher levels of vitamin D in their blood have a lower risk of colon cancer. Vitamin D helps protect the cells in the colon against changes that can lead to cancer. People can get vitamin D through sunlight and through their diets.

Some people get vitamin D by taking vitamins. Good food sources include dairy products and breakfast cereals that are fortified with vitamin D and fatty fish like salmon and tuna.

Vitamin D is also important for bone health and can help prevent osteoporosis (bone loss).

Risk Perception Questions

Your Beliefs About Cancer Prevention Activities
In the following set of questions, we ask you to rate how effective each of the activities are in preventing each type of cancer. Please note that each block of questions refers to a specific cancer type.
For each cancer prevention strategy, please state how effective you believe it is in reducing the risk of **colon** cancer.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Not effective. Does not reduce cancer risk.</th>
<th>Somewhat effective. Reduces risk by 10 to 30% of average.</th>
<th>Moderately effective. Reduces risk by 30 to 60% of average.</th>
<th>Very effective. Reduces risk by 60 to 80% of average.</th>
<th>Extremely effective. Reduces risk by more than 80% of average.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being physically active for more than 30 minutes most days or at least 3 hours per week.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Taking a multivitamin on most days.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Taking a baby aspirin every day for at least 15 years.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Taking a vitamin D supplement on most days.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Having a colonoscopy or sigmoidoscopy every 10 years after 50 years of age.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Maintaining a healthy weight.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

For each cancer prevention strategy, please state how effective you believe it is in reducing the risk of **bladder** cancer.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Not effective. Does not reduce cancer risk.</th>
<th>Somewhat effective. Reduces risk by 10 to 30% of average.</th>
<th>Moderately effective. Reduces risk by 30 to 60% of average.</th>
<th>Very effective. Reduces risk by 60 to 80% of average.</th>
<th>Extremely effective. Reduces risk by more than 80% of average.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being physically active for more than 30 minutes most days or at least 3 hours per week.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Taking a multivitamin with B-vitamins, Vitamin D and calcium on most days.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Taking a baby aspirin every day for at least 15 years.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Maintaining a healthy weight.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

For each cancer prevention strategy, please state how effective you believe it is in reducing the risk of **prostate** cancer.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Not effective. Does not reduce cancer risk.</th>
<th>Somewhat effective. Reduces risk by 10 to 30% of average.</th>
<th>Moderately effective. Reduces risk by 30 to 60% of average.</th>
<th>Very effective. Reduces risk by 60 to 80% of average.</th>
<th>Extremely effective. Reduces risk by more than 80% of average.</th>
</tr>
</thead>
</table>
Your Beliefs About Things that Increase the Risk of Cancer

In the following set of questions, we ask you to rate how much each of these conditions and activities increase the risk of cancer. Please note that each block of questions refers to a specific cancer type.

For each statement, please select the category that best reflects how much you believe it elevates the risk of **colon cancer**.

<table>
<thead>
<tr>
<th>Condition</th>
<th>No risk increase.</th>
<th>Small risk increase. Risk is higher, but less than double the average risk.</th>
<th>Moderate increase. Risk is 2 to 4 times the average risk.</th>
<th>Large increase. Risk is 4 to 8 times the average risk.</th>
<th>Very large increase. Risk is more than 8 times the average risk.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinking more than one alcoholic drink per day (one can of beer, a glass of wine, or a shot of hard liquor.)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Eating three or more servings of red meat per week.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Being overweight or obese.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Having a parent or sibling with colon cancer.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Being taller than 5'10&quot;.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Having chronic irritable bowel disease for 10 years or more.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Eating fewer than 3 servings of dairy each day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Smoking between 1 - 15 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Smoking 15 - 25 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Smoking more than 25 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

For each statement, please select the category that best reflects how much you believe it elevates the risk of **bladder cancer**.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Small risk increase. Risk is higher, but less than double the average risk.</th>
<th>Moderate</th>
<th>Large increase. Risk is 2 to 4 times the average risk.</th>
<th>Very large increase. Risk is more than 8 times the average risk.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking between 1 and 15 cigarettes per day</td>
<td>higher, but less than double the average risk</td>
<td>increase. Risk is 2 to 4 times the average risk</td>
<td>Risk is 4 to 8 times the average risk</td>
<td>increase. Risk is more than 8 times the average risk</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Smoking 15 to 25 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Smoking more than 25 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Working in a place where you are exposed to asbestos, radon, arsenic, or chromium for 5 - 20 years.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Drinking water with arsenic concentrations above the federal limit of 10 parts per billion.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Having a family history of bladder cancer.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

For each statement, please select the category that best reflects how much you believe it elevates the risk of **prostate cancer**.

<table>
<thead>
<tr>
<th>Drinking more than one alcoholic drink per day without taking a multivitamin (one can of beer, a glass of wine, or a shot of hard liquor.)</th>
<th>No risk increase.</th>
<th>Small risk increase. Risk is higher, but less than double the average risk</th>
<th>Moderate increase. Risk is 2 to 4 times the average risk</th>
<th>Large increase. Risk is 4 to 8 times the average risk</th>
<th>Very large increase. Risk is more than 8 times the average risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Having a father or brother with prostate cancer.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Being of African-American descent.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Being taller than 5'10&quot;.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Gaining between 22 and 44 lbs. since age 18.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Gaining 45 lbs. or more since age 18.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Having a vasectomy.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Smoking 1 - 14 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Smoking 15 - 25 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Smoking more than 25 cigarettes per day.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

The next three questions will ask you about what you believe your own personal risk is of getting these cancers in your lifetime compared to the typical man your age. Remember, we want your opinion so there are no wrong answers. Base your answers on what you currently believe the risk factors are for each cancer.
I think my risk of getting **colon cancer** in my lifetime compared to the typical man my age is:

- Zero. There is no chance of me getting colon cancer
- Very much below average (80 to 95% less than average)
- Much below average (80 to 50% less than average)
- Below average (50 to 10% less than average)
- Average
- Above average (0.1 to 2 times the average)
- Much above average (2 to 5 times the average)
- Very much above average (5 times or more higher than average)

I think my risk of getting **bladder cancer** in my lifetime compared to the typical man my age is:

- Zero. There is no chance of me getting bladder cancer
- Very much below average (80 to 95% less than average)
- Much below average (80 to 50% less than average)
- Below average (50 to 10% less than average)
- Average
- Above average (0.1 to 2 times the average)
- Much above average (2 to 5 times the average)
- Very much above average (5 times or more higher than average)

I think my risk of getting **prostate cancer** in my lifetime compared to the typical man my age is:

- Zero. There is no chance of me getting prostate cancer
- Very much below average (80 to 95% less than average)
- Much below average (80 to 50% less than average)
- Below average (50 to 10% less than average)
- Average
- Above average (0.1 to 2 times the average)
- Much above average (2 to 5 times the average)
- Very much above average (5 times or more higher than average)

**sMPL questions**
Hypothetical Health Risk and Preferred Treatments

Assume you have been diagnosed with a disease that will certainly be fatal in a year without treatment. There are two treatments (Treatment A and Treatment B), but neither is usually effective 100% of the time. Assume the costs of the treatment are the same, and neither treatment has side effects.

Each of the questions that follow will ask you whether you prefer Treatment A or Treatment B. For treatment A, the effectiveness does not change.

Treatment A: 30% of the time you will live an additional 8 years in good health, but 70% of the time you will live only 2 years in good health.

In Treatment B, the odds and outcomes are different from Treatment A.

Treatment B: 90% of the time you will only live 1 more year, but 10% of the time you will live much longer. How much longer you will live will be different for every question.

In the following set of questions, your task is to determine whether you prefer Treatment A or Treatment B. Click the radio button for the preferred treatment. Remember:

THERE ARE NO WRONG ANSWERS. THE QUESTION IS ABOUT WHAT TREATMENT YOU PREFER.

Select the treatment option you prefer.

Treatment A 30% chance of living 8 years and 70% chance of living 2 years
Treatment B 10% chance of living 13.5 years and 90% chance of living 1 more year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

Treatment A 30% chance of living 8 years and 70% chance of living 2 years
Treatment B 10% chance of living 15 years or 90% chance of living 1 year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

Treatment A 30% chance of living 8 years and 70% chance of living 2 years
Treatment B 10% chance of living 16.5 years or 90% chance of living 1 year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

Treatment A 30% chance of living 8 years and 70% chance of living 2 years
Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

- Treatment A 30% chance of living 8 years and 70% chance of living 2 years
- Treatment B 10% chance of living 18 years or 90% chance of living 1 year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

- Treatment A 30% chance of living 8 years and 70% chance of living 2 years
- Treatment B 10% chance of living 25 years or 90% chance of living 1 year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

- Treatment A 30% chance of living 8 years and 70% chance of living 2 years
- Treatment B 10% chance of living 30 years or 90% chance of living 1 year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

- Treatment A 30% chance of living 8 years and 70% chance of living 2 years
- Treatment B 10% chance of living 37 years or 90% chance of living 1 year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

- Treatment A 30% chance of living 8 years and 70% chance of living 2 years
- Treatment B 10% chance of living 44 years or 90% chance of living 1 year

Select the treatment option you prefer. Note that the treatment B outcome has changed (see underlined number in treatment B).

- Treatment A 30% chance of living 8 years and 70% chance of living 2 years
- Treatment B 10% chance of living 60 years or 90% chance of living 1 year

Thank you. The next section will ask you about your opinions about the effectiveness of cancer prevention activities and how other activities increase cancer risks. Please click the arrow to move to the next section.
## Your Health History

In the following section, we ask specific questions about you and your family's health history. Remember, all the information is entirely confidential and the results will be completely confidential.

### Have you ever had any type of cancer? (except non-melanoma skin cancer)
- Yes
- No

### What is your height?
- 5 foot 4 inches or less
- 5 foot 5 inches
- 5 foot 6 inches
- 5 foot 7 inches
- 5 foot 8 inches
- 5 foot 9 inches
- 5 foot 10 inches
- 5 foot 11 inches
- 6 foot 0 inches
- 6 foot 1 inch
- 6 foot 2 inches
- 6 foot 3 inches
- 6 foot 4 inches
- 6 foot 5 inches
- 6 foot 6 inches
- 6 foot 7 inches
- 6 foot 8 inches
- 6 foot 9 inches
- 6 foot 10 inches or more

### How much did you weigh at age 18 (in lbs)?
- 125 lbs or less
- 126 - 130 lbs
- 131 - 135 lbs
- 136 - 140 lbs
- 141 - 145 lbs
- 146 - 150 lbs
- 151 - 155 lbs
- 156 - 160 lbs
- 161 - 165 lbs
- 166 - 170 lbs
- 171 - 175 lbs
- 176 - 180 lbs
- 181 - 185 lbs
- 186 - 190 lbs
- 191 - 195 lbs
- 196 - 200 lbs
- 201 - 210 lbs
- 211 - 220 lbs
- 221 - 240 lbs
- 241 - 260 lbs
- 261 - 280 lbs
- 281 - 300 lbs
- 301 - 320 lbs
- More than 320 lbs

### What is your current weight (in lbs)?
- 125 lbs or less
- 126 - 130 lbs
- 131 - 135 lbs
- 136 - 140 lbs
- 141 - 145 lbs
- 146 - 150 lbs
- 151 - 155 lbs
- 156 - 160 lbs
- 161 - 165 lbs
- 166 - 170 lbs
- 171 - 175 lbs
- 176 - 180 lbs
- 181 - 185 lbs
- 186 - 190 lbs
- 191 - 195 lbs
- 196 - 200 lbs
- 200 - 210 lbs
- 211 - 220 lbs
- 221 - 240 lbs
- 241 - 260 lbs
- 261 - 280 lbs
- 281 - 300 lbs
- 301 - 320 lbs
- More than 320 lbs

Please check the correct box to indicate whether family members have had the cancer types given below. None of my immediate
<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Father</th>
<th>Mother</th>
<th>Brother or sister</th>
<th>Cancer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Cancer</td>
<td></td>
<td></td>
<td></td>
<td>Colon Cancer</td>
</tr>
<tr>
<td>Colon Cancer</td>
<td></td>
<td></td>
<td></td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>Bladder Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How much did you weigh when you were born?
- Under 8 1/2 pounds
- 8 1/2 pounds or more

Have you had chronic inflammatory bowel disease for 10 years or more? This includes Crohn’s disease and ulcerative colitis.
- Yes
- No

Health Related Activities

Health Related Activities

Do you walk or do other moderate activity for at least 30 minutes on most days, or at least 3 hours per week?
- Yes
- No

How many servings of alcohol do you have on a typical day? One serving is a can of beer, a glass of wine, or a shot of hard liquor.
- 0
- 1
- 2
- 3 or more

Have you had a vasectomy?
- Yes
- No
Have you taken aspirin every day for 15 years or more?

Yes

No

Do you eat 3 or more servings of red meat per week? 1 serving is 4 ounces—about the size of a deck of cards.

Yes

No

Do you eat 3 or more servings (1/2 cup) of tomatoes per week?

Yes

No

More Health Related Activities

How many servings of milk or dairy products do you have most days? One serving is one cup of milk, a cup of yogurt, or about 1 1/2 oz of cheese.

Fewer than 1.

1-2.

3 or more.

Do you take calcium supplements most days?

Yes

No

Do you take a multivitamin 4 or more days per week?

Yes

No

Do you take vitamin D supplements or calcium + vitamin D supplements on most days (apart from a standard multivitamin)?

Yes

No
Within the last 10 years, have you had a colonoscopy, sigmoidoscopy, virtual colonoscopy, or a fecal occult blood test?
Yes
No

Select the best response that describes your smoking history
I currently am a smoker
I used to smoke, but I quit.
I have never been a smoker.

How much do you smoke?
14 or fewer cigarettes per day
15-25 cigarettes per day
more than 25 cigarettes per day

Have you ever worked in the production of rubber or aluminum or were you exposed to aromatic amines for 5 years or more?
Yes
No

When working with the substances, did you use protective gear? Protective gear includes respirators, eye protection, gloves and boots.
Yes
No

What’s the total amount of time you worked with these substances (without protective gear)?
Less than 5 years
5 to 20 years
more than 20 years

Do you get your drinking water from a private well?
Yes
No

Health Insurance Questions
Health Insurance

Describe your current health insurance.

- I do not have any health insurance.
- I buy my own health insurance and my employer does not contribute.
- I have health insurance provided by my employer and no other insurance.
- I have health insurance provided by my employer and I also purchase additional insurance with my own money.
- I have Medicare insurance and employer-provided health insurance.
- I have Medicare insurance and no employer-provided insurance.
- I have Medicaid insurance.
- I have insurance through the Veteran’s administration.

What is your co-payment for each office visit to a general practitioners such as internists or nurse practitioners? (Not for emergency room or hospital stays or for specialists. These may be different).

- I have a zero copay.
- My copay is $25 or less
- My copay is $25 to $50
- My copay is more than $50.

What is your annual deductible?

- I do not have an annual deductible.
- I have an annual deductible between $1 and $1,000
- I have an annual deductible between $1,001 and $3,000
- I have an annual deductible between $3,001 and $5,000
- My annual deductible is more than $5,000.

Thank you for completing this section. The next section will ask you a hypothetical scenario for an insurance that would cover cancer costs. Note: these questions are for scientific investigation only. All information is confidential and we are not trying to sell you anything.

Cancer Insurance Description

How Much Would You Be Willing to Pay for Cancer Insurance?

Assume that there is an insurance policy available that will cover any and all costs related to the covered cancers. The
costs including diagnostic testing, office visits for specialists, hospital stays, treatment costs including chemotherapy and radiation, as well as FDA approved experimental treatments. There are no copays or deductibles and you would be able to choose your own doctors and hospitals. **Please assume that your current insurance will not cover these cancers and that you will have to pay all of the costs yourself if you get any of these cancers.**

**WTP questions**

**Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer?** Remember that we asked you to assume, for the purposes of answering this question, that if you don’t purchase the insurance that you will be responsible for all costs if you get any of these cancers.

- **No.** I would not pay this much for this insurance.
- **Yes.** I would be willing to pay $5 per month for this insurance.

**Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer?** Remember that we asked you to assume, for the purposes of answering this question, that if you don’t purchase the insurance that you will be responsible for all costs if you get any of these cancers.

- **No.** I would not pay this much for this insurance.
- **Yes.** I would be willing to pay $15 per month for this insurance.

**Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer?** Remember that we asked you to assume, for the purposes of answering this question, that if you don’t purchase the insurance that you will be responsible for all costs if you get any of these cancers.

- **No.** I would not pay this much for this insurance.
- **Yes.** I would be willing to pay $25 per month for this insurance.

**Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer?** Remember that we asked you to assume, for the purposes of answering this question, that if you don’t purchase the insurance that you will be responsible for all costs if you get any of these cancers.

- **No.** I would not pay this much for this insurance.
- **Yes.** I would be willing to pay $35 per month for this insurance.

**Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer?** Remember that we asked you to assume, for the purposes of answering this question, that if you don’t purchase the insurance that you will be responsible for all costs if you get any of these cancers.

- **No.** I would not pay this much for this insurance.
- **Yes.** I would be willing to pay $45 per month for this insurance.

**Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer?** Remember that we asked you to assume, for the purposes of answering this question, that if you don’t
purchase the insurance that you will be responsible for all costs if you get any of these cancers.
- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $55 per month for this insurance.

Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer? Remember that we asked you to assume, for the purposes of answering this question, that if you don't purchase the insurance that you will be responsible for all costs if you get any of these cancers.
- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $65 per month for this insurance.

Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer? Remember that we asked you to assume, for the purposes of answering this question, that if you don't purchase the insurance that you will be responsible for all costs if you get any of these cancers.
- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $75 per month for this insurance.

Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer? Remember that we asked you to assume, for the purposes of answering this question, that if you don't purchase the insurance that you will be responsible for all costs if you get any of these cancers.
- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $85 per month for this insurance.

Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer? Remember that we asked you to assume, for the purposes of answering this question, that if you don't purchase the insurance that you will be responsible for all costs if you get any of these cancers.
- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $95 per month for this insurance.

Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer? Remember that we asked you to assume, for the purposes of answering this question, that if you don't purchase the insurance that you will be responsible for all costs if you get any of these cancers.
- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $105 per month for this insurance.

Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer? Remember that we asked you to assume, for the purposes of answering this question, that if you don't purchase the insurance that you will be responsible for all costs if you get any of these cancers.
- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $120 per month for this insurance.
Would you be willing to pay the amount below for this cancer insurance that covers only colon, prostate, and bladder cancer? Remember that we asked you to assume, for the purposes of answering this question, that if you don’t purchase the insurance that you will be responsible for all costs if you get any of these cancers.

- No, I would not pay this much for this insurance.
- Yes, I would be willing to pay $135 per month for this insurance.

Demographic Questions

Demographic Questions
The following page asks you about demographic characteristics of your household. We realize some of these questions are quite personal. We promise to keep all of your information completely confidential.

What is your annual household income including your income, any income from a spouse, and any others who contribute to your household budget?

- $9,999 or less
- $10,000 to $19,999
- $20,000 to $29,999
- $30,000 to $39,999
- $40,000 to $49,999
- $50,000 to $59,999
- $60,000 to $69,999
- $70,000 to $79,999
- $80,000 to $89,999
- $90,000 to $99,999
- $100,000 to $109,999
- $110,000 to $119,999
- $120,000 to $129,999
- $130,000 to $139,999
- $140,000 to $159,999
- $160,000 to $189,999
- $190,000 to $209,999
- $210,000 or more

What race or ethnic group best describes you?

- Black/African American
- Hispanic
- Asian/South Asian/Pacific Islander
- White/Nonhispanic
- Native American
- Mostly Jewish Ethnicity
- Other

What is your age in years?

What is your level of education?

- Didn’t finish high school
- High school graduate
- Some college
- 2 year college degree
- 4 year college degree
- Professional or other postgraduate degree

Marital Status
Cognitive ability questions

The next page includes four questions that measure verbal, numerical, and abstract reasoning. You will have 1 1/2 minutes to answer the questions. Please click the arrow to start this section.

These page timer metrics will not be displayed to the recipient.
First Click: 0 seconds.
Last Click: 0 seconds.
Page Submit: 0 seconds.
Click Count: 0 clicks.

0130

The girl sings in a choir. All singers in the choir play a musical instrument. The girl plays a musical instrument. Assuming the first 2 statements are true, the final statement is:

- True
- Neither True nor False
- False

PRESENT RESERVE-Do these words:
- Have the same meaning
- Have contradictory meaning
- Mean neither the same nor opposite
- I don't know

Look at the numbers below. What number comes next?
8, 4, 2, 1, 1/2, 1/4, ?
- 4
- 1/2
- 1/8
- 1/4
- 1/16
- I don't know
The hours of daylight and darkness in SEPTEMBER are nearest equal to the hours of daylight and darkness in:

- May
- March
- April
- November

The next page includes three questions that measure verbal, numerical, and abstract reasoning. You will have 2 1/2 minutes to answer the questions. Please click the arrow to start this section.

These page timer metrics will not be displayed to the recipient.
First Click: 0 seconds.
Last Click: 0 seconds.
Page Submit: 0 seconds.
Click Count: 0 clicks.

Minutes is to time as centimeters is to...

- Measure
- Long
- Length
- Second
- Meter
- I don't know

A train travels 20 feet in 1/5 seconds. At this exact speed, how many feet will it travel in three seconds?

- 100 feet
- 200 feet
- 300 feet
- 275 feet
- I don't know

Three individuals form a partnership and agree to divide all the profits equally. X invests $9,000, Y invests $7,000 and Z...
○ $1,600
○ $560
○ $400
○ $800
○ I don't know
References


CURRICULUM VITAE

David S. Hales

EDUCATION
Massachusetts Institute of Technology, S.B. in Physics, 1994
National University, B.S. in Computer Science, 1997
University of Hawaii-Manoa, MBA, 2003
University of Nevada-Las Vegas, M.A. in Economics, 2015 (Anticipated)

GRADUATE-LEVEL ECONOMICS COURSEWORK
Public Finance, UNLV (Grade: A)
Mathematical Economics, UNLV (Grade: A)
Econometrics, UNLV (Grades: A, A)
Microeconomics, UNLV (Grade: A)
Economic Theory & Policy, UNLV (Grade: A)
Economic Research Seminar, UNLV (Grades: A, A)
Macroeconomics (Spring 2015)

GRADUATE-LEVEL MATHEMATICS COURSEWORK
Advanced Matrix Theory & Applications, UNLV (Grade: A)
Introduction to Real Analysis, UNLV (Grade: A)
Introduction to Math Statistics, UNLV (Grade: A)

SALIENT UNDERGRADUATE MATHEMATICS COURSEWORK
Introduction to Complex Analysis, Brigham Young University (Grade: A-)
Introduction to Probability and Statistics, MIT (Grade: A)

ACADEMIC PAPER
Mary Riddel and David Hales, “Risk Misperceptions and Selection in Insurance Markets: An Application to Demand for Cancer Insurance,” presented at the March 2014 Risk, Perception, and Response Conference at the Harvard School of Public Health, and at the December 2014 Behavioral Insurance Workshop at Ludwig-Maximilians University of Munich. This paper has been submitted for publication consideration.
TEACHING EXPERIENCE

Teaching Assistant, Public Finance, UNLV (Fall 2013)
Supplemental Instructor, Intermediate Microeconomics, UNLV (Spring 2014)
Teaching Assistant, Business Economics, UNLV (Fall 2014)

PRINCIPAL EMPLOYMENT

10/2011 – 7/2015 Board Chairman and CEO, Global Fidelity Corp.
6/2008 – 9/2011 Board Member and CFO, Global Fidelity Corp.

OTHER APPOINTMENTS

9/2009 – Managing Director, Ohana Matters Foundation
10/2009 – Steering Committee Member, U.S. Senator Brian Schatz (D-HI)
11/2013 – Military Affairs Advisor, Mayor John Lee, North Las Vegas
12/2008 – Reserve Officer (Lieutenant Colonel), U.S. Air Force Reserves
4/2004 – 7/2015 Board Member, Pipeline Micro, Inc.
10/2013 – 10/2014 Advisory Board Member, College of Southern Nevada
4/2009 – 12/2010 Board Member, Hawaii Science and Technology Council
11/2008 – 6/2010 Operations Committee Member, Hawaii Angels Investor Group