Discussion Paper Series – CRC TR 224

Discussion Paper No. 069
Project A 02

Risky Lifestyle Choices of Women With Breast Cancer

Chloé Michel [1]
Michelle Sovinsky [2]
Steven Stern [3]

March 2021
(First version : February 2019)

[1] Swiss Re
[2] University of Mannheim and CEPR
[3] Stony Brook University

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Risky Lifestyle Choices of Women with Breast Cancer

Chloé Michel, Michelle Sovinsky, and Steven Stern[1]

February 26, 2021

Abstract: Using data from the Panel Study of Income Dynamics on breast cancer diagnosis and lifestyle choices, we estimate how being diagnosed influences smoking, drinking, and exercising habits for more than 8,000 women over the period 1999 to 2011. Controlling for unobserved heterogeneity, persistence in potentially addictive behaviors, and correlation across behaviors, we find that the impact of a diagnosis has a different effect on smoking, drinking, and exercising behavior. Furthermore, the impact depends upon the recency of the diagnosis. Recently diagnosed women exercise and smoke less but do not change their drinking habits relative to healthy women. Our approach provides insight into what extent women who are faced with negative information about life expectancy take this into consideration when deciding to engage in risky behaviors that might further affect their survival in a significant way.

Keywords: breast cancer, risky health behavior, health economics

JEL Codes: I12, J16, C35

---

[1] Michel is at Swiss Re; Sovinsky is at the University of Mannheim and CEPR; Stern is at Stony Brook University. Corresponding author is Sovinsky (University of Mannheim, Department of Economics, L7, 3-5 68161 Mannheim, Germany; msovinsky@econ.uni-mannheim.de).We thank Janet Currie and Michael Darden for helpful comments. We gratefully acknowledge support from the Swiss National Science Foundation Grant #130333. The collection of data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and the National Science Foundation under award number 1157698. Sovinsky acknowledges support from the European Research Council Grant #725081 FORENSICS and from the CRC Transregio Grant 224 (A02),

1 Introduction

About 13% of US women will develop breast cancer at some point during their life, and world-wide incidence is rising.[2] There are a number of genetic and demographic factors linked to breast cancer risk, in addition several lifestyle habits are associated with incidence including weight gain, fat intake, and level of physical activity, while others have been inconsistently linked with the disease including alcohol consumption and cigarette smoking.[3] Whether to engage in physical activity, drink alcohol, or smoke are choices associated with how to live.[4] Therefore, understanding lifestyle decisions made by diagnosed women can provide useful information about the tradeoffs women are willing to make between participating in unhealthy habits and increasing one's life expectancy.

Individuals with a breast cancer diagnosis are a particularly informative group to learn about the value of engaging in risky behaviors. First, as noted earlier, lifestyle habits are associated with breast cancer incidence. Second, breast cancer is a cancer with one of the highest survival rates - nearly 90% of patients survive in the first 5 years.[5] Third, it is also one of the cancers with the highest recurrence rates. Almost 30% of patients with breast cancer who are free of the disease after initial treatment(s) have a recurrence during follow-up (Saphner, et. al., 1996).[6] These facts together suggest that choices made among these individuals can be used to inform us about the value of risky behaviors because (i) behaviors influence incidence, (ii) there is an incentive to change behavior to combat recurrence, and (iii) the sample size is large enough as there are a substantial number people diagnosed and who survive more than 5 years post-diagnosis.

---

[2] Surveillance Epidemiology and End Results survey (Gloeckler Ries et al., 2007).

[3] For example, see Berry et al. (2005), Demark-Wahnefried et al. (2000), Singletary and Gapstur (2001), Pinto et al. (2002), and Blanchard et al. (2004).

[4] There are numerous papers that examine risky behaviors such as smoking, drinking, and exercise/obesity. See for example, Perreira and Sloan (2001), Khwaja et al. (2006), and Klijs et al. (2011).

[5] According to the National Cancer Institute, the top-5 cancers with the highest five year survival rates are prostate (98.6%), thyroid (98.2%), testicular (95.1%), lip (91%), breast (89.7%). This does not include breast cancer in situ, which has a 100% survival rate as it lacks the invasive nature of cancer. https://www.cancer.gov/about-cancer/understanding/statistics accessed October 21, 2019.

[6] Those cancers with the highest rates of recurrence include glioblastoma (recurs in nearly all patients), ovarian cancer (85%), non-Hodgkin lymphoma PTCL (75%), soft tissue sarcoma (50%), bladder (50%), pancreatic (36% to 46%), non-Hodgkin lymphoma DLBCL (30% to 40%), breast (30%). Source: https://www.cancertherapyadvisor.com/home/tools/fact-sheets/cancer-recurrence-statistics/ accessed October 21, 2019.

We examine the impact a breast cancer diagnosis has on engaging in (potentially addictive) risky behaviors over time. The Panel Study of Income Dynamics contains rich longitudinal information on the timing of breast cancer diagnosis and lifestyle choices that we use to estimate the model. This approach illustrates to what extent women who are faced with negative information about life expectancy take this into consideration when deciding to engage in risky behaviors that might further affect their survival in a significant way.

Our parameter estimates indicate that breast cancer diagnosis impacts lifestyle choices. However, the impact of diagnosis has a different effect on smoking, drinking, and exercising behavior, and the impact also depends upon the recency of the diagnosis. We find that women who had a diagnosis within the last five years exercise less and smoke less but do not change their drinking habits relative to healthy women. These changes in behavior are not always consistent with information provided to the public on breast cancer risk factors.[7] However, these choices may be rationalized when one considers the overall value of life where lifestyle choices increase the utility from living.

There are numerous studies in the economics and medical literatures that examine issues associated with breast cancer.[8] However, there are relatively few that consider the relationship with lifestyle choices,[9] and, to the best of our knowledge, ours is the first paper to examine changes in behavior while controlling for persistence in lifestyle choices. Among those papers that examine lifestyle choices among breast cancer survivors, Bellizi et. al., (2005) conduct a descriptive analysis of the prevalence of health behaviors (smoking, alcohol use, physical activity, and cancer screening) of cancer survivors by age, time since diagnosis, and cancer site using data from the National Health Interview Survey. They find that cancer survivors are more likely to meet the recommendations for physical activity and cancer screening compared with noncancer controls. However, they do not find any evidence of different behavior among survivors with respect to smoking and alcohol consumption. We complement and add to the previous studies in a number of ways. First, we use a large,

[7] There is a related literature on how publicly available information and guidelines impact behavior. See for example, Hu et al. (1995), Ippolito and Mathios (1995), and Jacobson and Kadiyala (2017).

[8] These include studies on cancer mortality (e.g., Cutler, 2008), investment in research (e.g., Budish, Roin, and Williams, 2015), mammography screening (e.g., Bitler and Carpenter, 2016; Jacobson and Kadiyala, 2017), costs of treatment (e.g., Einav, Finkelstein, and Williams, 2016), and insurance coverage (e.g., Decker, 2005; Jerome-D'Emilia et. al., 2010). We do not provide a comprehensive review of this vast literature.

[9] These include Blanchard et al. (2004), Chan et. al., (2016), Demark-Wahnefried et al. (2000), Pinto et. al., (2002), Braithwaite et al. (2012) who focus on smoking, Ibrahim and Al-Homaidh (2011) who focus on physical activity, and Singletary and Gapstur (2001) who focus on alcohol consumption.

nationally representative sample that includes women diagnosed with breast cancer. Second, we examine changes in lifestyle behaviors over time where we allow for persistence in behavior.

We proceed as follows in the rest of the paper. In the next section, we discuss the data. In section 3, we present a framework that links breast cancer diagnosis and risky lifestyle choices. We discuss the estimation methodology in section 4 and the estimation results in section 5. We conclude in section 6.

2 Data

Our research uses data from the Panel Study of Income Dynamics (PSID). The PSID is a longitudinal study that started in 1968 and now includes more than 22,000 individuals from over 9,000 households in the United States. One person per family, designated as the "head," is interviewed biennially and answers questions about the individuals of the household.[10] Every wave contains information about employment, income, education, wealth, marriage, childbearing, and various other topics. We choose to use the PSID data set because of its longitudinal structure which allows us to follow the same individuals and their corresponding behaviors across time. Further, these data are collected not only for breast cancer patients but also for persons without a history of cancer. This allows us to make comparisons between breast cancer patients and healthy individuals.

We use data from seven waves from 1999, when cancer outcomes were first recorded, until 2011. We retain respondents who are aged 15 and older and are female because breast cancer almost exclusively affects women. After dropping individuals who have missing information on demographics,[11] breast cancer condition, or (lagged) lifestyle behaviors, we have a sample of 8,028 women and 34,109 person-years. Some of these women have missing information on one lifestyle behavior but not another. For our analysis on each behavior, we drop only those observations with missing values for questions related to those behaviors. So, for smoking habits, this subsample includes 8,019 women and 33,947 person-years; for exercise

[10] The head of the household provides answers for questions related to his or her spouse. The literature has shown that spouses have very precise perceptions of the time spent by the other spouse on different activities (Stern, 2003). Similarly, it has been shown (see, for example, Kolonel et al., 1977; Mejia et al., 2017) that spouses provide complete information for various lifestyle behaviors of their spouse such as smoking and drinking behaviors.

[11] This corresponds to missing values for age, race, education level, or income.

it includes $8,009$ women and $33,851$ person-years, while for drinking it is smaller (for reasons we discuss momentarily) and includes $7,175$ women and $18,082$ person-years.

Table 1 reports demographic summary statistics, and Table 2 reports health behaviors summary statistics for our sample. The PSID was initially designed to study the dynamics of

Variable	**Mean**	**Std. Dev.**
Age	46.278	15.598
White	0.582	0.493
Black	0.305	0.461
Married	0.638	0.481
Employed	0.632	0.482
Has Children	0.870	0.337
Highest Education Degree:		
High School	0.418	0.493
University	0.325	0.468
Post Graduate	0.090	0.286
Taxable Income:		
$<$ \$ 20,000	0.183	0.387
$>$ \$ 50,000	0.228	0.420
Diagnosed with:		
Cancer	0.094	0.292
Breast Cancer	0.023	0.151
Number of Person-Years:	34109	

Table 1: Demographics Summary Statistics

income and poverty. The oversampling of families who were poor in the late 1960s resulted in a substantial subsample of blacks (PSID, 2013). In our sample, we also have a large proportion of black respondents (30%). One of our main interests is the health status of our respondents and, in particular, their cancer status. As can be seen in Table 1, 9.4% of the sample have been diagnosed with cancer and 2.3% with breast cancer, which matches the proportion of cancers that are breast cancers reported in the national breast cancer statistics of the American Cancer Society (2007). As can be seen in Table 2, approximately 54% of our respondents ever drink alcoholic beverages, which is slightly below the national average of 55% as reported by the National Center for Health Statistics (Schoenborn and Adams, 2010) for the period 2005-2007.[12] As the survey questions concerning alcohol consumption

[12] They report the proportion of current drinkers, which refer to adults who have had at least 12 drinks in their lifetime and at least one drink in the past year. Looking at these numbers disaggregated by race, we find in our sample that 61% of the white respondents and 43% of the blacks ever drink alcoholic beverages.

Variable	Mean	Std. Dev.	#Person-Years
Smoking Status			33967
Current Smoker	0.176	0.381	
Cigarette Consumption			5987
Smokes 1 to 9 cig/day	0.334	0.472	
Smokes 10 to 19 cig/day	0.363	0.481	
Smokes 20 or more cig/day	0.303	0.460	
Alcohol			18082
Drinks Alcohol	0.543	0.498	
Frequency of Alcohol Consumption[1]			9814
Less than 1 drink per month	0.291	0.454	
One drink per month	0.205	0.404	
Several drinks per month	0.159	0.366	
One drink per week	0.167	0.373	
Several drinks per week	0.137	0.344	
Drinks everyday	0.041	0.198	
Exercise			33851
Never	0.172	0.377	
1 or 2 times/week	0.179	0.384	
3 to 6 times/week	0.305	0.460	
7 times/week	0.308	0.462	
8 to 14 times/week	0.016	0.126	
More than 14 times/week	0.020	0.140	

[1] This includes only waves 2005, 2007, 2009, and 2011.

Table 2: Health Behaviors Summary Statistics

were not consistently worded across waves, we report statistics only for the last four waves (2005, 2007, 2009, and 2011).[13]

Table 3 presents details about respondents with breast cancer. Individuals in the sample

Variable	Mean	Std. Dev.	# Person-Years
Years Since BC Diagnosis	11.255	12.204	1472
Age at BC Diagnosis	51.394	14.734	
Currently[1]			
Cured	0.759	0.428	
In Remission	0.145	0.352	
In Treatment	0.096	0.295	

[1] These questions are asked starting only in 2005.

Table 3: Descriptive Statistics for Individuals with Breast Cancer

This also matches the numbers in the National Center for Health Statistics which report 59% and 40% of current drinkers for whites and blacks respectively.

[13] For the first three waves (1999, 2001, and 2003), people are asked how many drinks they have *on average* per day: "In the last year, on average, how often did you have any alcohol to drink? Would you say, less than one a month, about once a month, several times a month, about once a week, several times a week, or every day?" For the last four waves , the categories were changed and the questions about daily consumption referred to *days when respondents drink*: "In the last year, on the days you drank, about how many drinks did you have?" In later regressions, we also use data only from years 2005, 2007, 2009, and 2011 when looking at alcohol behaviors.

Variable	*Proportion of Breast Cancer Diagnoses* Mean	Std. Dev.	Person-Years	p-value[1]
Race				0.000***
White	0.023	0.166	19838	
Black	0.018	0.133	10413	
Age				0.000***
Younger than 30	0.000	0.020	5187	
Between 30 and 59	0.017	0.130	22657	
60 and older	0.065	0.247	6265	
Family Composition				0.000***
Have Children	0.025	0.156	28581	
Childless	0.016	0.124	5528	
Education				0.045*
No High School Diploma	0.026	0.161	5736	
High School	0.024	0.155	14249	
Associate or Bachelor	0.021	0.143	11069	
>Bachelor	0.020	0.140	3055	
Family Income				0.000***
<\$20,000	0.034	0.182	6246	
≥\$20,000 & <\$50,000	0.019	0.136	7782	
>\$50,000	0.022	0.146	20081	

[1] The reported p-values are from multivariate tests on equal means.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Proportion of Breast Cancer Diagnoses Disaggregated by Demographics

responded to the following question: "Has a doctor ever told you that you have or had cancer or a malignant tumor?" If the respondent answered "yes," follow-up questions were asked regarding the type of cancer and the stage. The majority of our respondents are "cured," while approximately 9% are in treatment.[14] As can be seen in Table 3, the sample average age for a breast cancer diagnosis is approximately 51.

In Table 4, we report prevalence of breast cancer diagnosis by demographic groups. The proportion of respondents having breast cancer is larger among whites than among individuals of other races. This is in line with national statistics, which indicate that white women have the highest probability of getting diagnosed with breast cancer (American Cancer So-

[14] Cancer is considered as "cured" when doctors cannot detect cancer five years after diagnosis (American Cancer Society, 2006). Questions about whether the respondent is currently in treatment, in remission, or has been cured are asked only starting in 2005. The sample size is therefore smaller.

	Proportion of Breast Cancer Diagnoses			
Variable	**Mean**	**Std. Dev.**	**Person-Years**	**p-value**[1]
Smoking Status				0.000***
Smoker	0.015	0.122	5987	
Non-Smoker	0.025	0.157	27980	
Cigarette Consumption				0.037*
Smokes 1 to 9 cig/day	0.012	0.107	1998	
Smokes 10 to 19 cig/day	0.014	0.119	2174	
Smokes 20 or more cig/day	0.020	0.139	1815	
Alcohol				0.000***
Drinks Alcohol	0.020	0.140	9814	
Never Drinks Alcohol	0.030	0.169	8268	
Frequency of Alcohol Consumption[2]				0.058
Less than 1 drink per month	0.019	0.136	2853	
One drink per month	0.019	0.138	2009	
Several drinks per month	0.015	0.120	1563	
One drink per week	0.020	0.140	1640	
Several drinks per week	0.027	0.161	1346	
Drinks everyday	0.030	0.170	403	
Exercise				0.000***
Never	0.037	0.188	5817	
1 or 2 times/week	0.021	0.144	6070	
3 to 6 times/week	0.021	0.143	10318	
7 times/week	0.021	0.142	10423	
8 to 14 times/week	0.013	0.113	546	
More than 14 times/week	0.024	0.152	677	

[1] The reported p-values are from multivariate tests on equal means.

[2] This considers only waves 2005, 2007, 2009 and 2011.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5: Proportion of Breast Cancer Diagnoses Disaggregated by Health Behaviors

ciety, 2005). However, black women are less likely to use diagnostic services, and, when they are diagnosed, it is typically at a later stage (American Cancer Society Statistics, 2003).[15]

Next, we examine relationships between risky health behaviors and breast cancer prevalence. Table 5 displays breast cancer diagnosis among individuals with differing smoking, drinking, and exercise habits.[16] Among smokers, breast cancer prevalence is the highest for respondents who smoke more than 19 cigarettes per day. Regarding alcohol consumption

[15] Diagnosis requires going to the doctor, and women without adequate insurance are going to be less likely to go to the doctor.

[16] There are some women in the sample who have breast cancer but have not yet been diagnosed. However, given that the woman does not know she has breast cancer, this will not influence her actions and hence will not impact our results.

behaviors, prevalence is lower in the group of respondents who drink alcohol. Among those who drink, breast cancer prevalence is highest among those women who drink more than one drink per week.

The bottom panel of Table 5 presents statistics for physical activity. The respondents were asked about weekly exercise frequency for heavy and light workouts. Specifically, they were asked, "How often do you participate in vigorous/light physical activity or sports?" A problem with this wording is that there is no information about the measure of time spent by a person doing physical activity.[17] In our analysis, we first aggregate light and heavy physical activities into one variable called "exercise." Second, we define the following six categories: no exercise (neither light nor heavy), exercise $1-2$ times a week, exercise $3-6$ times a week, exercise 7 times a week, exercise $8-14$ times a week, and exercise more than 14 times a week.[18] The proportion of breast cancer patients is the largest among people who never exercise. The main point that emerges from Table 5 is that breast cancer incidence differs with the degree that an individual engages in lifestyle behaviors. In the next sections, we examine this in more detail.

3 Econometric Specification

In our framework, a woman (indexed by i) makes a lifestyle choice (indexed by l) in each period (indexed by t), where the lifestyle behaviors may be influenced by breast cancer diagnosis. The lifestyle choices concern how much to smoke, how much to consume alcohol, and how much to engage in physical activity. Let y^*_{ilt} be a latent variable measuring the continuous quantity of lifestyle activity l chosen by individual i at time t. Specifically, the baseline model is given by

$$y^*_{ilt} = 1\left(y^*_{ilt-1} > 0\right)\alpha_l + X_{it}\eta_l + b_{it}\delta_l + \mu_{il} + \varepsilon_{ilt}. \tag{1}$$

Lifestyle choices exhibit persistence, which may be due to addiction (such as smoking and drinking alcohol) or habit persistence (such as exercise). Therefore, we allow individual

[17] Specifically, heavy exercise refers to "heavy housework, aerobics, running, swimming, bicycling or similar activity that causes heavy sweating or large increases in breathing or heart rate" (PSID, 2005). Light exercise includes "walking, dancing, gardening, golfing, bowling or similar activity that causes only light sweating or slight to moderate increases in breathing or heart rate" (PSID, 2005). See the discussion in Berniell et al. (2013).

[18] Some persons report extreme values which could indicate some misunderstanding of the question.

i's lifestyle choices at time t to depend on whether she participated in that behavior in the immediate past $1(y^*_{ilt-1} > 0)$. Exogenous, possibly time-varying individual demographic variables (X_{it}) that are likely to influence lifestyle choices include i's age, her marital status, whether she has children, her income, and her education level.

There may be heterogeneity that we do not observe in the data that influences choices and has a persistent nature. Unobserved heterogeneity likely to influence lifestyle choices includes a person/behavior-specific random effect μ_{il}, which captures things such as taste for alcohol or dislike of exercise, and an idiosyncratic effect ε_{ilt}.

Whether a woman has been diagnosed with breast cancer (captured by dummy b_{it}) may impact her decision to engage in risky behaviors, for example, if she feels that these behaviors may reduce her longevity more severely than prior to the breast cancer diagnosis. To the extent that smoking, drinking, or exercise are risk factors for getting breast cancer, one may be concerned that b_{it} is a function of prior choices. In effect, causation may run in both directions. We address issues of endogeneity and unobserved heterogeneity using Wooldridge (2002) fixed effects techniques that we describe momentarily. Finally, we need to include an initial value of the risky decisions at time $t = 0$. These are likely to be endogenous, and we follow previous literature (à la Heckman, 1981) to control for endogenous initial conditions. We specify the initial period values as

$$y^*_{il0} = C_i\varsigma_l + X_{i0}\eta_l + b_{i0}\delta_l + \mu_{il} + \varepsilon_{il0}, \tag{2}$$

$$\varepsilon_{il0} \sim iidN(0, \sigma^2_\varepsilon),$$

where C_i is a set of variables affecting only initial choices. For smoking behaviors, these include age when i started smoking.[19] Unfortunately, the PSID does not contain any information on the age at which respondents started drinking or exercising. For these lifestyle choices, we include the level of drinking or exercising behavior observed in the first period of the data as C_i. In this approach, there may be a concern about the value of μ_{il}. One possibility is to treat it as random, which would imply that μ_{il} and y^*_{il0} are independent. However, μ_{il} and y^*_{il0} may not be independent, so we follow Wooldridge (2002) that builds on Chamberlain (1984) and specify the construction of the fixed effect conditional on the initial condition as

[19] For those who do not smoke, the initial condition is set to zero. This is an innocuous normalization because we control for those who have never smoked. A separate concern is that age started smoking might be endogenous. However, most of the literature (e.g., Wooldridge, 2005) ignores this issue.

$$\begin{aligned} \mu_{il} &= \pi_0 + \pi_1 y^*_{il0} + \overline{X}_i \pi_2 + v_{il}, \\ v_{il} &\sim N(0, \sigma_v^2) \end{aligned} \quad (3)$$

where $\overline{X}_i$ denotes the mean over time of the explanatory variables (excluding the year fixed effects). As discussed in Wooldridge (2005), the random component of the fixed effect then can be integrated out to yield the likelihood function of the random effects probit model with time-t, observation-i explanatory variables: $(X_{it}, y_{il,t-1}, ...y_{il0}, \overline{X}_i)$ (we define y_{ilt} momentarily).

4 Estimation Methodology

We begin by estimating three models corresponding to the lifestyle activities separately. Then we allow for correlation across smoking, drinking, and exercise behaviors by estimating all decisions jointly. However, due to data restrictions that we discuss in section (2), some of these behaviors are recorded only for a subset of the data. We estimate the parameters of our model by a straightforward dynamic ordered probit estimation methodology. In the context of our framework, using equation (1), we define

$$Y_{ilt}(\mu_{il}) = 1\left(y^*_{ilt-1} > 0\right)\alpha_l + X_{it}\eta_l + b_{it}\delta_l + \mu_{il}$$

as the deterministic part of y^*_{ilt} after conditioning on μ_{il} for $t \geq 1$. Then using equation (2), we can define

$$Y_{il0}(\mu_{il}) = C_i \varsigma_l + X_{i0}\eta_l + b_{i0}\delta_l + \mu_{il};$$

where, using equation (3), we define

$$\overline{\mu}_{il} = \pi_0 + \pi_1 y^*_{il0} + \overline{X}_i \pi_2.$$

There is only one small complication: each behavior is reported in the data as a bracketed variable. To account for this, we define $m = 1$ as not participating in the activity and let the quantity of the activity increase as m increases with

$$y_{ilt} = m \text{ iff } \kappa_{lm} \leq y^*_{ilt} < \kappa_{lm+1}, \;\; m = 1, 2, .., M_l, \quad (4)$$

where κ_{lm} are cutoff points to be estimated; this is akin to a dynamic ordered probit model.[20]

The vector of parameters to estimate for model l is $\theta_l = (\alpha_l, \eta_l, \delta_l, \varsigma_l, \pi_0, \pi_1, \pi_2, \sigma_\varepsilon, \sigma_v, \kappa_l)$, and the log likelihood contribution for lifestyle choice l individual i is

$$L_{il} = \log \int \left[\prod_{t=1}^{T} \left(\prod_{m=1}^{M_l} \Delta_{ilm} \left(\mu_{il}\right)^{1(y_{ilt}=m)} \right) \left(\prod_{m=1}^{M_l} \Delta_{ilm} \left(\frac{\mu_{il}}{\sigma_\varepsilon} \right)^{1(y_{il0}=m)} \right) \frac{1}{\sigma_v} \phi \left(\frac{\mu_{il} - \overline{\mu}_{il}}{\sigma_v} \right) d\mu_{il} \right]$$

where

$$\Delta_{ilm} \left(\frac{\mu_{il}}{\sigma} \right) = \Phi \left(\frac{\kappa_{lm+1} - Y_{ilt}\left(\mu_{il}\right)}{\sigma} \right) - \Phi \left(\frac{\kappa_{lm} - Y_{ilt}\left(\mu_{il}\right)}{\sigma} \right)$$

for $\sigma = 1$ or σ_ε. The log likelihood function $L_l = \sum_i L_{il}$ can be evaluated using a quadrature method (Butler and Moffitt, 1982).[21] Finally, to allow for correlation across risky lifestyle choices, we estimate a (dynamic) multivariate ordered probit model (see Appendix A for details).

5 Parameter Estimates

Our parameter estimates indicate that breast cancer diagnosis and recency of diagnosis impacts lifestyle choices. However, the impact of diagnosis has a different effect on smoking, drinking, and exercising, and the impact also depends upon the recency of the diagnosis. We start by discussing the impact of breast cancer diagnosis on smoking behavior. We consider four categories of daily smoking intensity: (i) does not smoke; (ii) smokes but fewer than 10 cigarettes a day; (iii) smokes between 10 and 19 cigarettes a day; or (iv) smokes more than 20 daily (which is more than a pack of cigarettes). Table 6 presents random-effects ordered probit estimates where the explanatory variables include smoking behavior in the previous year, demographics, as well as breast cancer variables.

[20] Assume without loss of generality that $\kappa_{l1} = -\infty$, $\kappa_{l2} = 0$, and $\kappa_{lM_l+1} = \infty$.

[21] We use antithetic acceleration in simuluation. Geweke (1988) shows that the loss of precision in simulation is of the order $1/N$ (where N is the number of observations), which does not require an adjustment to the asymptotic covariance matrix.

Dependent Variable: Ordered Variable for Number of Cigarettes Smoked	(1)	(2)	(3)	(4)
Lagged Behavior				
Smoker Last Period	2.432***	1.646***	2.433***	1.648***
	(0.0298)	(0.0450)	(0.0298)	(0.0450)
Breast Cancer Variables				
Diagnosed with Breast Cancer	-0.0957	-0.150		
	(0.0984)	(0.142)		
Recent Breast Cancer Diagnosis			-0.289**	-0.326*
			(0.140)	(0.171)
Other Controls				
Aged in 30s, 40s, or 50s	-0.0279	0.172***	-0.0272	0.172***
	(0.0329)	(0.0509)	(0.0329)	(0.0509)
Aged 60 or Older	-0.519***	-0.160	-0.520***	-0.164
	(0.0513)	(0.103)	(0.0511)	(0.103)
White	0.474***	0.701***	0.473***	0.700***
	(0.0527)	(0.0842)	(0.0527)	(0.0841)
Black	0.0807	0.196**	0.0798	0.195**
	(0.0550)	(0.0882)	(0.0549)	(0.0881)
Married	-0.227***	-0.259***	-0.227***	-0.259***
	(0.0296)	(0.0394)	(0.0296)	(0.0393)
Have Children	0.00395	-0.0321	0.00392	-0.0321
	(0.0417)	(0.0632)	(0.0417)	(0.0631)
Highest Education is High School	-0.241***	-0.352***	-0.241***	-0.352***
	(0.0361)	(0.0516)	(0.0361)	(0.0516)
Highest Education is University Degree	-0.505***	-0.700***	-0.505***	-0.700***
	(0.0422)	(0.0605)	(0.0422)	(0.0604)
Highest Education is Post Graduate	-0.896***	-1.268***	-0.895***	-1.266***
	(0.0748)	(0.107)	(0.0748)	(0.107)
Income Less than 20K	0.102***	0.109***	0.102***	0.109***
	(0.0346)	(0.0418)	(0.0346)	(0.0418)
Income Between 20 and 50K	0.0953***	0.106***	0.0949***	0.105***
	(0.0310)	(0.0375)	(0.0310)	(0.0375)
Initial Conditions Included	no	yes	no	yes
Number of Observations	33,967	33,942	33,967	33,942
Number of Individuals	8,019	8,010	8,019	8,010

Notes: Standard errors in parentheses. * indicates significance at 10% level; ** at 5%; and *** at 1%.
All regressions include cut-off points, individual heterogeneity variance and year fixed effects.
The initial conditions specifications include the mean over time of all time varying regressors.

Table 6: Random-Effects Ordered Probit Regressions for Smoking

The signs and significance of the control variables are intuitive and consistent with results from other studies. First, past smokers are more likely to be current smokers, and the significant positive effect persists after controlling for unobserved heterogeneity (in columns 2 and 4). Our finding is consistent with numerous studies that have shown that smoking exhibits true state dependence (i.e., the effect is significant after controlling for unobserved heterogeneity). Our results indicate that white women smoke more than black women (see Schoenborn and Adams, 2010). We also find that married women smoke less than those who are not married as do women with a higher education relative to other education categories. Finally, we find that individuals with lower incomes (under $\$50,000$) smoke more than higher-income women.

The first two columns indicate that whether an individual has been diagnosed with breast cancer has no significant impact on smoking behavior conditional on past behavior and demographic variables. However, as columns (3) and (4) show, if the woman was diagnosed with breast cancer less than five years ago, she will significantly decrease her smoking behavior

(-0.289) with this effect being robust to including initial conditions (column 4, -0.326). The differential impact of the time of diagnosis on smoking behavior could arise from a few sources. First, the individual may react to a diagnosis by curbing unhealthy habits such as smoking, but this effect may deteriorate over time as the individual survives past the initial stages. Second, the woman may be undergoing treatment which makes smoking more difficult in the short term due to lack of energy, for example.

Table 7 presents the results of a random-effects ordered probit regression for number of alcoholic drinks, where the dependent variable is ordered according to: (i) a non-drinker, (ii) a woman who drinks at most once a week on average, and (iii) a woman who drinks more than once a week on average. As with smoking, our results indicate that past drinking behavior is a positive significant indicator of current drinking behavior, and this effect remains after controlling for initial conditions in columns (2) and (4). The other control variables imply that women aged 60 or older drink less than younger women and that white women drink more than black women. In addition, we find that married women drink less often as do those with children. Drinking more often is more likely among those with higher education relative to other groups and among those with a larger income. In contrast to smoking behaviors, women do not change their alcohol consumption after a breast cancer diagnosis regardless of when the diagnosis was made.

Dependent Variable: Ordered Variable for Number of Alcoholic Drinks	(1)	(2)	(3)	(4)
Lagged Behavior				
Number of Drinks Last Period	0.220***	0.0513***	0.220***	0.0513***
	(0.0111)	(0.0119)	(0.0111)	(0.0119)
Breast Cancer Variables				
Diagnosed with Breast Cancer	-0.0791	-0.101		
	(0.136)	(0.142)		
Recent Breast Cancer Diagnosis			-0.0112	-0.0874
			(0.180)	(0.186)
Other Controls				
Aged in 30s, 40s, or 50s	0.0125	0.104**	0.0116	0.103**
	(0.0481)	(0.0498)	(0.0480)	(0.0498)
Aged 60 or Older	-0.394***	-0.163**	-0.398***	-0.167**
	(0.0673)	(0.0699)	(0.0670)	(0.0696)
White	0.849***	0.716***	0.848***	0.715***
	(0.0737)	(0.0764)	(0.0736)	(0.0764)
Black	0.137*	0.148*	0.136*	0.148*
	(0.0777)	(0.0811)	(0.0777)	(0.0811)
Married	-0.192***	-0.151***	-0.192***	-0.151***
	(0.0430)	(0.0444)	(0.0430)	(0.0444)
Have Children	-0.474***	-0.391***	-0.474***	-0.391***
	(0.0591)	(0.0610)	(0.0591)	(0.0610)
Highest Education is High School	0.465***	0.467***	0.465***	0.467***
	(0.0605)	(0.0629)	(0.0605)	(0.0629)
Highest Education is University Degree	0.804***	0.802***	0.804***	0.802***
	(0.0644)	(0.0669)	(0.0644)	(0.0669)
Highest Education is Post Graduate	1.008***	1.017***	1.009***	1.017***
	(0.0818)	(0.0849)	(0.0818)	(0.0849)
Income Less than 20K	-0.0721	-0.0563	-0.0722	-0.0564
	(0.0452)	(0.0467)	(0.0452)	(0.0467)
Income Between 20 and 50K	-0.0615	-0.0613	-0.0614	-0.0612
	(0.0388)	(0.0401)	(0.0388)	(0.0401)
Initial Conditions Included	no	yes	no	yes
Number of Observations	18,082	18,036	18,082	18,036
Number of Individuals	7,175	7,147	7,175	7,147

Notes: Standard errors in parentheses. * indicates significance at 10% level; ** at 5%; and *** at 1%.
All regressions include cut-off points, individual heterogeneity variance and year fixed effects.
The initial conditions specifications include the mean over time of all time varying regressors.

Table 7: Random-Effects Ordered Probit Regressions for Alcohol Consumption

We present the results of the random-effects ordered probit for exercise frequency in Table 8. Exercise frequency is based on the number of exercise sessions per week as discussed in section (2) where the categories are whether one participates in exercise (i) no times, (ii) $1-2$ times, (iii) $3-6$ times, (iv), 7 times, (v) $8-14$ times, or (vi) more than 14 times, per week. The control variables indicate that, the older the woman is, the less physical activity she participates in. The results also show that being white is associated with higher levels of physical activity. Our findings also indicate that married women engage in more physical activity relative to non-married women. Furthermore, the higher the level of education the woman has, the more she engages in weekly physical activity. Finally, individuals with income less than $\$20,000$ engage in less exercise relative to individuals with income between $\$20,000$ and $\$50,000$.

Dependent Variable: Ordered Variable for Exercise Frequency	(1)	(2)	(3)	(4)
Lagged Behavior				
Exercise Frequency Last Period	0.173***	0.126***	0.174***	0.126***
	(0.00727)	(0.00731)	(0.00727)	(0.00731)
Breast Cancer Variables				
Diagnosed with Breast Cancer	-0.144***	-0.167***		
	(0.0505)	(0.0516)		
Recent Breast Cancer Diagnosis			-0.138**	-0.156**
			(0.0698)	(0.0705)
Other Controls				
Aged in 30s, 40s, or 50s	-0.134***	-0.147***	-0.135***	-0.148***
	(0.0194)	(0.0197)	(0.0194)	(0.0197)
Aged 60 or Older	-0.381***	-0.392***	-0.386***	-0.399***
	(0.0257)	(0.0261)	(0.0256)	(0.0260)
White	0.155***	0.134***	0.154***	0.133***
	(0.0257)	(0.0262)	(0.0256)	(0.0262)
Black	-0.0624**	-0.0664**	-0.0629**	-0.0670**
	(0.0274)	(0.0280)	(0.0274)	(0.0280)
Married	0.0547***	0.0565***	0.0546***	0.0564***
	(0.0167)	(0.0170)	(0.0167)	(0.0170)
Have Children	0.0172	0.0174	0.0169	0.0171
	(0.0231)	(0.0235)	(0.0231)	(0.0235)
Highest Education is High School	0.113***	0.103***	0.113***	0.103***
	(0.0223)	(0.0227)	(0.0223)	(0.0227)
Highest Education is University Degree	0.163***	0.154***	0.163***	0.154***
	(0.0239)	(0.0243)	(0.0239)	(0.0243)
Highest Education is Post Graduate	0.198***	0.199***	0.198***	0.199***
	(0.0322)	(0.0328)	(0.0322)	(0.0328)
Income Less than 20K	0.0482**	0.0393*	0.0479**	0.0389*
	(0.0199)	(0.0201)	(0.0199)	(0.0201)
Income Between 20 and 50K	0.0639***	0.0572***	0.0638***	0.0572***
	(0.0174)	(0.0176)	(0.0174)	(0.0176)
Initial Conditions Included	no	yes	no	yes
Number of Observations	33,851	33,851	33,851	33,851
Number of Individuals	8,009	8,009	8,009	8,009

Notes: Standard errors in parentheses. * indicates significance at 10% level; ** at 5%; and *** at 1%.
All regressions include cut-off points, individual heterogeneity variance and year fixed effects.
The initial conditions specifications include the mean over time of all time varying regressors.

Table 8: Random Effects Ordered Probit Regressions for Exercising

As the results in columns (1) and (2) show, a diagnosis of breast cancer significantly impacts the amount of exercise in a negative way. Perhaps this result is not so surprising given that women often undergo treatment after a breast cancer diagnosis that can weaken them and make it more difficult to engage in extra physical activity. The results in columns (3) and (4) show that women also decrease their amount of physical activity after a recent diagnosis.

In addition, we find that the impact of breast cancer diagnosis for each risky behaviors remains after we control for other changes in a woman's life in the last year. Specifically, we include changes in marital status (i.e., getting married or divorced), changes in health status (i.e., moving into a state of poorer (self-reported) health), and changes in employment status (i.e., become employed or losing a job). We re-estimated the specifications from Tables 6-8 with additional covariates measuring changes. The results show no significant changes in the impact of a breast cancer diagnosis and the impact of a recent diagnosis on behaviors. We report the parameter estimates in Appendix B.

	(1) smoke	(1) drink	(1) exercise	(2) smoke	(2) drink	(2) exercise
Lagged Behavior						
Smoker Last Period	2.534***			2.535***		
	(0.0221)			(0.0222)		
Number of Drinks Last Period		0.438***			0.438***	
		(0.00614)			(0.00614)	
Exercise Frequency Last Period			0.291***			0.291***
			(0.00531)			(0.00530)
Breast Cancer Variables						
Diagnosed with Breast Cancer	-0.0640	-0.0251	-0.105***			
	(0.0728)	(0.0539)	(0.0387)			
Recent Breast Cancer Diagnosis				-0.274**	0.0142	-0.102*
				(0.113)	(0.0858)	(0.0611)
Other Controls						
Aged in 30s, 40s, or 50s	0.0322	0.0324	-0.122***	0.0334	0.0319	-0.123***
	(0.0265)	(0.0231)	(0.0166)	(0.0265)	(0.0231)	(0.0166)
Aged 60 or Older	-0.271***	-0.126***	-0.329***	-0.271***	-0.128***	-0.334***
	(0.0371)	(0.0292)	(0.0209)	(0.0369)	(0.0291)	(0.0208)
White	0.319***	0.368***	0.111***	0.319***	0.367***	0.110***
	(0.0363)	(0.0274)	(0.0192)	(0.0363)	(0.0274)	(0.0192)
Black	-0.0326	0.0973***	-0.0481**	-0.0335	0.0971***	-0.0485**
	(0.0384)	(0.0293)	(0.0206)	(0.0384)	(0.0293)	(0.0206)
Married	-0.135***	-0.0476**	0.0427***	-0.135***	-0.0475**	0.0427***
	(0.0222)	(0.0190)	(0.0134)	(0.0222)	(0.0190)	(0.0134)
Have Children	0.0520*	-0.171***	0.0173	0.0523*	-0.171***	0.0172
	(0.0302)	(0.0239)	(0.0176)	(0.0302)	(0.0239)	(0.0176)
Highest Education is High School	-0.167***	0.310***	0.0814***	-0.167***	0.310***	0.0812***
	(0.0261)	(0.0254)	(0.0172)	(0.0261)	(0.0254)	(0.0172)
Highest Education is University Degree	-0.364***	0.579***	0.108***	-0.364***	0.579***	0.108***
	(0.0299)	(0.0266)	(0.0183)	(0.0299)	(0.0266)	(0.0183)
Highest Education is Post Graduate	-0.604***	0.738***	0.135***	-0.604***	0.738***	0.135***
	(0.0519)	(0.0346)	(0.0248)	(0.0519)	(0.0346)	(0.0248)
Income Less than 20K	0.0761***	0.0710***	-0.00975	0.0763***	0.0709***	-0.0101
	(0.0272)	(0.0231)	(0.0167)	(0.0272)	(0.0231)	(0.0167)
Income Between 20 and 50K	0.0600**	-0.00619	0.00633	0.0599**	-0.00620	0.00626
	(0.0242)	(0.0202)	(0.0146)	(0.0242)	(0.0202)	(0.0146)
Covariance Terms						
Smoking and Drinking	0.0478***			0.0479***		
	(0.0141)			(0.0141)		
Smoking and Exercise	-0.0227**			-0.0227**		
	(0.0100)			(0.0100)		
Drinking and Exercise	0.0549***			0.0549***		
	(0.00852)			(0.00852)		

Notes: Standard errors in parentheses. * indicates significance at 10% level; ** at 5%; and *** at 1%. All regressions include cut-off points. Number of observations is 34,109.

Table 9: Multivariate Ordered Probit Regressions

It may be the case that decisions to smoke, drink, or exercise are correlated with eachother. Table 9 presents the estimates from dynamic multivariate ordered probit regressions which allows for this correlation. These regressions use information on all behaviors over all periods during which they are available, hence the sample size is somewhat smaller. The first specification includes information on whether an individual was diagnosed with breast cancer and the second includes only a recent diagnosis. We continue to find a significant impact of past behaviors, which is not surprising. The results indicate that indeed there is correlation across behaviors (as evidenced by the significant covariance terms). However, the estimates of the impact of a breast cancer diagnosis remain and are consistent with those from Tables 6-8. Namely, a recent breast cancer diagnosis results in less smoking and exercise, but does not impact alcohol consumption. In summary, our results paint a consistent picture of how a breast cancer diagnosis influences a woman's decision to engage in risky lifestyle choices.

6 Conclusions

According to the National Breast Cancer Foundation, one in eight US women are impacted by breast cancer.[22] We use longitudinal data from the PSID, starting from 1999 to 2011, to examine to what extent women who are diagnosed with breast cancer change their (potentially risky) lifestyle choices. After controlling for unobserved heterogeneity, persistence in potentially addictive behaviors, and correlation across behaviors, we find that women who were recently diagnosed with breast cancer smoke less. In contrast to smoking behavior, women do not change their alcohol consumption after a breast cancer diagnosis regardless of when the diagnosis was made. Furthermore, a diagnosis of breast cancer significantly impacts the amount of exercise in a negative way. Perhaps this latter result is not so surprising given that women often undergo treatment after a breast cancer diagnosis that can weaken them and make it more difficult to engage in extra physical activity. Our findings provide insight into what extent women who are faced with negative information about life expectancy take this into consideration when deciding to engage in risky behaviors that might further affect their survival in a significant way.

[22] See https://www.nationalbreastcancer.org/what-is-breast-cancer
Referenced on October 28, 2016

Appendix

A Multivariate Ordered Probit Details

To allow for correlation across choices in the error structure, we estimate a dynamic multivariate ordered probit model where

$$\varepsilon_{it} = \begin{pmatrix} \varepsilon_{i1t} \\ \varepsilon_{i2t} \\ \varepsilon_{i3t} \end{pmatrix} \sim iidN\left[0, \Omega\right]$$

with

$$\Omega = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix},$$

the vector of parameters is augmented to include $(\rho_{12}, \rho_{13}, \rho_{23})$, and the log likelihood contribution for individual i is

$$L_i = \log \int \left[\prod_{t=1}^{T} \left(\prod_{\overrightarrow{m}} \Delta_{i\overrightarrow{m}} \left(\mu_i\right)^{1\left(y_{it} = \overrightarrow{m}\right)} \right) \left(\prod_{\overrightarrow{m}} \Delta_{i\overrightarrow{m}} \left(\mu_i\right)^{1\left(y_{i0} = \overrightarrow{m}\right)} \right) \prod_{l=1}^{3} \left(\frac{1}{\sigma_v} \phi \left(\frac{\mu_{il} - \overline{\mu}_{il}}{\sigma_v} \right) d\mu_{il} \right) \right]$$

where $\overrightarrow{m} = (m_1, m_2, m_3)$ is the vector of discrete choices made for each risky behavior, $\mu_i = (\mu_{i1}, \mu_{i2}, \mu_{i3})$, $y_{it} = (y_{i1t}, y_{i2t}, y_{i3t})$, and

$$\Delta_{i\overrightarrow{m}} \left(\frac{\mu_i}{\sigma} \right) = \Pr\left[\kappa_{lm_l} - Y_{ilt}\left(\mu_{il}\right) \leq \varepsilon_{ilt} \leq \kappa_{lm_l+1} - Y_{ilt}\left(\mu_{il}\right), l = 1, 2, 3 \right]$$

which is the trivariate normal density with covariance matrix Ω integrated over the rectangle with sides defined by the limits of ε_{ilt}.

B Results for Additional Specifications

	Risky Activity					
	smoke	smoke	drink	drink	exercise	exercise
Breast Cancer Variables						
Diagnosed with Breast Cancer	-0.0997		-0.0678		-0.130***	
	(0.0985)		(0.136)		(0.0503)	
Recent Breast Cancer Diagnosis		-0.288**		0.00192		-0.128*
		(0.140)		(0.180)		(0.0697)
Lagged Behavior						
Smoker Last Period	2.428***	2.430***				
	(0.0298)	(0.0298)				
Number of Drinks Last Period			0.219***	0.219***		
			(0.0111)	(0.0111)		
Exercise Frequency Last Period					0.174***	0.174***
					(0.00727)	(0.00727)
Changes in Covariates						
Change in Marital Status	0.150***	0.149***	-0.00638	-0.00637	0.0229	0.0230
	(0.0387)	(0.0387)	(0.0506)	(0.0506)	(0.0241)	(0.0241)
Moved into Poor Health	0.138**	0.137**	-0.363***	-0.364***	-0.548***	-0.549***
	(0.0682)	(0.0682)	(0.0936)	(0.0936)	(0.0426)	(0.0425)
Change in Employment Status	0.0662**	0.0664**	0.0529	0.0530	0.0501***	0.0504***
	(0.0293)	(0.0293)	(0.0371)	(0.0371)	(0.0172)	(0.0172)
Other Controls						
Aged in 30s, 40s, or 50s	-0.0128	-0.0121	0.0173	0.0165	-0.124***	-0.124***
	(0.0331)	(0.0331)	(0.0482)	(0.0481)	(0.0195)	(0.0195)
Aged 60 or Older	-0.494***	-0.495***	-0.379***	-0.383***	-0.349***	-0.354***
	(0.0518)	(0.0516)	(0.0676)	(0.0673)	(0.0259)	(0.0258)
White	0.480***	0.480***	0.848***	0.847***	0.151***	0.150***
	(0.0527)	(0.0527)	(0.0736)	(0.0736)	(0.0255)	(0.0255)
Black	0.0896	0.0886	0.133*	0.133*	-0.0671**	-0.0676**
	(0.0549)	(0.0549)	(0.0777)	(0.0777)	(0.0273)	(0.0273)
Married	-0.219***	-0.219***	-0.196***	-0.196***	0.0510***	0.0509***
	(0.0296)	(0.0296)	(0.0431)	(0.0431)	(0.0167)	(0.0167)
Have Children	-0.00353	-0.00357	-0.475***	-0.476***	0.0145	0.0143
	(0.0417)	(0.0417)	(0.0591)	(0.0591)	(0.0229)	(0.0229)
Highest Education is High School	-0.237***	-0.237***	0.458***	0.458***	0.0988***	0.0986***
	(0.0361)	(0.0361)	(0.0605)	(0.0605)	(0.0222)	(0.0222)
Highest Education is University Degree	-0.504***	-0.504***	0.795***	0.795***	0.147***	0.147***
	(0.0423)	(0.0423)	(0.0644)	(0.0644)	(0.0238)	(0.0238)
Highest Education is Post Graduate	-0.896***	-0.895***	0.997***	0.998***	0.178***	0.178***
	(0.0749)	(0.0748)	(0.0819)	(0.0818)	(0.0321)	(0.0321)
Income Less than 20K	0.0867**	0.0867**	-0.0763*	-0.0763*	0.0429**	0.0425**
	(0.0349)	(0.0349)	(0.0455)	(0.0455)	(0.0200)	(0.0200)
Income Between 20 and 50K	0.0898***	0.0894***	-0.0655*	-0.0654*	0.0592***	0.0591***
	(0.0310)	(0.0310)	(0.0389)	(0.0389)	(0.0174)	(0.0174)
Observations	33,967	33,967	18,082	18,082	33,851	33,851
Number of Individuals	8,019	8,019	7,175	7,175	8,009	8,009

Notes: Standard errors in parentheses. * indicates significance at 10% level; ** at 5%; and *** at 1%. All regressions include cut-off points.

Table B1: Regressions including changes in covariates

References

American Cancer Society Statistics (2003). *Cancer Facts & Figures for African Americans 2003–2004.* Atlanta, GA: ACS, 2003. http://www.cancer.org/downloads/STT/861403.pdf.

American Cancer Society (2005). *Breast Cancer Facts and Figures 2005-2006.* Atlanta, GA: American Cancer Society.

American Cancer Society (2006). *Cancer Facts and Figures 2006.* Atlanta, GA: American Cancer Society.

American Cancer Society (2007). *Breast Cancer Facts and Figures 2007-2008.* Atlanta, GA: American Cancer Society.

Bellizzi, K., J. Rowland, D. Jeffery, and T. McNeel (2005). "Health Behaviors of Cancer Survivors: Examining Opportunities for Cancer Control Intervention." *Journal of Clinical Oncology.* 23(34): 8884-8892.

Berniell, Lucilia, Dolores De La Mata, and Nieves Valdes (2013). "Spillovers of Health Education at School on Parents' Physical Activity." *Health Economics.* 22(9): 1004-1020.

Berry, D., K. Cronin, S. Plevritis, D. Fryback, L.Clarke, M. Zelen, J. Mandelblatt, A. Yakovlev, Dik F. Habbema, and Eric J. Feuer (2005). "Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer." *New England Journal of Medicine.* 353(17): 1784–92.

Bitler M. and C. Carpenter (2016). "Health Insurance Mandates, Mammography, and Breast Cancer Diagnoses." *American Economic Journal: Economic Policy.* 8(3): 39-68.

Blanchard C., K. Stein, F. Baker, M. Dent, M. Denniston, K. Courney, and E. Nehl (2004). "Association Between Current Lifestyle Behaviors and Health-Related Quality of Life in Breast, Colorectal and Prostate Cancer Survivors." *Psychology and Health.* 19: 1-13.

Braithwaite, D., M. Izano, D. Moore, M. Kwan, M. Tammemagi, R. Hiatt, K. Kerlikowske, C. Kroenke, C. Sweeney, L. Habel, A. Castillo, E. Weltzien, and B. Caan (2012). "Smoking and Survival After Breast Cancer Diagnosis: A Prospective Observational Study and Systematic Review." *Breast Cancer Research and Treatment.* 136(2): 521-533.

Budish, E., B. Roin, and H. Williams (2015). "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials." *American Economic Review.* 105(7): 2044-2085.

Butler, J. and R. Moffitt (1982). "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model." *Econometrica.* 50(3): 761-764.

Centers for Disease Control and Prevention (CDCP). (2002) *Women and Smoking: A Report of the Surgeon General.* MMWR 51: 1-30.

Chamberlain, G. (1984). "Panel Data," in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Volume 2. Amsterdam: North Holland, 1247-1318.

Chan, T, B. Hamilton, and N. Papageorge (2016). "Health, Risky Behavior and the Value of Medical Innovation for Infectious Disease." *Review of Economic Studies.* 83(4): 1465–1510.

Cutler, David M. (2008). "Are We Finally Winning the War on Cancer?" *Journal of Economic Perspectives.* 22(4): 3–26.

Decker, S. (2005). "Medicare and the Health of Women with Breast Cancer." *The Journal of Human Resources.* 40(4): 948-68.

Demark-Wahnefried, W., B. Peterson, C. McBride, I. Lipkus, and E. Clipp (2000) "Current Health Behaviors and Readiness to Pursue Lifestyle Changes among Men and Women Diagnosed with Early Stage Prostate and Breast Carcinomas." *Cancer.* 88: 674-684.

Einav, L., A. Finkelstein, and H. Williams (2016). "Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments." *American Economic Journal: Economic Policy.* 8(1): 52-79.

Geweke, J. (1988). "Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference," *Journal of Econometrics*, 38(1-2): 73-89.

Gloeckler Ries, Lynn, John Young, Gretchen Keel, Milton Eisner, Yi Dan Lin, and Marie-Josephe Horner (2007). *SEER Survival Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988-2001, Patient and Tumor Characteristics.* National Cancer Institute, SEER Program, NIH Pub. No. 07-6215, Bethesda, MD, 2007 .

Heckman, J. (1981). "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process." in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications.* MIT Press: Cambridge, Mass.

Hu, T., H. Sung, and T. Keeler (1995). "The State Antismoking Campaign and the Industry Response: The Effects of Advertising on Cigarette Consumption in California." *American Economic Review.* 85(2): 85-90.

Ibrahim, E. and A. Al-Homaidh (2011). "Physical Activity and Survival After Breast Cancer Diagnosis: Meta-Analysis of Published Studies." *Medical Oncology.* 28: 753–765.

Ippolito, P. and A. Mathios (1995). "Information and Advertising: The Case of Fat Consumption in the United States." *American Economic Review.* 85(2): 91-95.

Jacobson, M. and S. Kadiyala (2017). "When Guidelines Conflict: A Case Study of Mammography Testing during the 1990s." *Women's Health Issues.* 27(6): 692-699.

Jerome-D'Emilia, B., E. Merwin, and S, Stern (2010). "Feasability of Using Technology to Disseminate Evidence to Rural Nurses and Improve Patient Outcomes," *Journal of Continuing Education in Nursing.*

Khwaja, A., F. Sloan, and S. Chung. (2006). "Learning about Individual Risk and the Decision to Smoke," *International Journal of Industrial Organization.* 24: 683-699.

Klijs, B., J. Mackenback, and A. Kunst (2011). "Obesity, Smoking, Alcohol Consumption and Years Lived with Disability: A Sullivan Life Table Approach." *BMC Public Health.* DOI: 10.1186/1471-2458-11-378.

Kolonel, L., T. Hirohata, and A. Nomura (1977). "Adequacy of Survey Data Collected from Substitute Respondents." *American Journal of Epidemiology.* 106(6): 476-484.

Mejia, R., S. Braun, L. Pena, S. Gregorich, and E. Perez-Stable (2017). "Validation of Non-Smoking Status by Spouse Following a Cessation Intervention." *Journal of Smoking Cessation.* 12(1): 38-42.

Panel Study of Income Dynamics (PSID) (2005). Public Release Family File. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

Panel Study of Income Dynamics (PSID) (2013). Main Interview User Manual. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

Perreira, K. and F. Sloan (2001). "Life Events and Alcohol Consumption among Mature Adults: A Longitudinal Analysis." *Journal of Studies on Alcohol.* 62(4): 501-508.

Pinto, B., N. Maruyama, M. Clark, D. Cruess, E. Park, and M. Roberts (2002). "Motivation to Modify Lifestyle Risk Behaviors in Women Treated for Breast Cancer." *Mayo Clinic Proceedings.* 77(2): 122-129.

Saphner, T., D. Tormey, and R. Gray (1996). "Annual Hazard Rates of Recurrence for Breast Cancer after Primary Therapy." *Journal of Clinical Oncology* 14: 2738–2746.

Schoenborn, C. and P. Adams (2010). "Health Behaviors of Adults: United States, 2005-2007." *Vital Health Statistics 10.* 245: 1-132.

Singletary, K. and S. Gapstur (2001). "Alcohol and Breast Cancer: Review of Epidemiologic and Experimental Evidence and Potential Mechanisms." *Journal of the American Medical Association.* 286(17): 2143-2151.

Stern, S. (2003). "Info from the National Survey of Families and Households on Asymmetric Information in Marriage." http://faculty.virginia.edu/stevenstern/resint/marriagestf/marriagesignals.html

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data.* MIT Press: Cambridge, Mass.

Wooldridge, J. (2005). "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Effects." *Journal of Applied Econometrics.* 20(1): 39-54.