UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2015

The Effect of Cancer on the Employment of Older Males: Attenuating Selection Bias using a High Risk Sample

David Candon, University College Dublin

WP15/07

March 2015

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

The Effect of Cancer on the Employment of Older Males: Attenuating Selection Bias using a High Risk Sample

David Candon*

School of Economics, University College Dublin, Ireland

ABSTRACT

Estimating the unbiased effect of health shocks on employment is an important topic in both health and labour economics. This is particularly relevant to cancer, where improvements in screening and treatments have led to increases in survival for nearly all types of cancer. In order to address the issue of selection bias, I estimate the effect of cancer on employment for a high-risk cancer sample, male workers over the age of 65, thus attenuating the impact of many cancer risk factors. This identification strategy balances the covariates between the cancer and the non-cancer groups in numerous tests. Respondents who are diagnosed with cancer are 13.2 percentage points less likely to work than their non-cancer counterparts. The results also appear insensitive to omitted confounders.

JEL classification: I10, I18, J21, J26

Keywords: Cancer; Employment; Labour market

*Corresponding author:
Mr D. Candon
D217 School of Economics
University College Dublin
Belfield
Dublin 4
Ireland
Tel: 00 353 1 7168505
Fax: 00 353 1 7168272
Email: david.candon@ucd.ie

1. INTRODUCTION

The goal of obtaining unbiased estimates of health shocks on employment is often confounded by the problem of selection. The most common example of this is workers who possess characteristics unobservable to the econometrician which simultaneously make them more likely to receive the health shock, but which also result in worse labour outcomes had they not received the shock. In this case, the effect of the health shock on employment is overestimated.

From a policy perspective, the benefits of obtaining unbiased estimates of this effect are clear. By knowing exactly how a health shock affects labour supply, the government can more accurately design programs to either rehabilitate the workers back into the labour force or to address their physical and psychological needs. This is particularly relevant to cancer, where improvements in screening and treatments have led to increases in survival for nearly all types of cancer. As of 2014, there are 14 million Americans alive who have been diagnosed with cancer at some point in their lives (American Cancer Society, 2014a).

In order to address the selection problem, I examine the effect of cancer diagnosis on male workers who are least 65 years of age. While many cancer risk factors have been identified, such as tobacco or alcohol use, old age remains the single biggest risk factor in developing cancer (American Society of Clinical Oncology, 2015a). By focusing the analysis on older males, who have never been diagnosed with cancer before, it attenuates the impact of these traditional risk factors on future cancer diagnosis. In addition, almost all other covariates are balanced between the cancer and non-cancer groups. Under these conditions, the issue of selection bias is reduced to the point of negligibility

Using data from the Health and Retirement Study, I demonstrate that cancer risk factors do little to predict future diagnosis with numerous tests. With bivariate OLS regression, I find that respondents who are diagnosed with cancer are 13.3 percentage points ($p<0.01$) less likely to work than their non-cancer counterparts. If I also control for 30 pre-treatment covariates, the effect falls to 13.2 percentage points ($p<0.01$). The robustness of the effect is examined by removing different subgroups groups. Depending on which subgroup is removed, the effect varies between 10 to 14 percentage points but the adjusted and unadjusted results never vary by more than one percentage point, again indicating that cancer diagnosis has little correlation with pre-treatment characteristics. I also demonstrate that an omitted confounding variable would have to be far more correlated with cancer diagnosis than any traditional cancer risk factor in order to be the true cause of the effect. Overall, the results present a convincing case for an unbiased effect of cancer on employment.

With regards to previous estimates of the effect of cancer on male labour supply in the U.S., Bradley et al. (2005) and Bradley et al. (2007) show that older men who have been diagnosed with prostate cancer are less likely to work, and work fewer hours, in the 6 months following prostate cancer diagnosis than healthy controls. However, this effect is absent at 12 and 18 after diagnosis. Looking at longer term outcomes, Short et al. (2008) shows that, between two to six years after diagnosis, there is no significant difference in employment for older male cancer survivors versus healthy workers though they are less likely to work full-time and they work fewer hours compared to control workers. The longer term impact for younger workers is similar, with Moran et al. (2011) showing that male survivors are less likely to be employed, less likely to work full-time, and work fewer hours than controls, in the two to six years following diagnosis.

All of these studies rely on the selection on observables assumption, either with OLS regression or propensity score matching, yet there is little reason to believe that this assumption is met. In all of these cases, there are significant differences between the treatment group and the control group across a wide range of demographic, health and employment related variables. While it is possible to account for observable differences, having the control and treatment groups differ wildly across such important characteristics does nothing to allay fears that the two groups could differ across important unobservable characteristics. The importance of this point cannot be overemphasised. As Heckman et al. (1997) point out, simple impositions of common support, reweighting of covariate densities, and using pre-treatment information on the outcome of interest cannot eliminate bias with non-experimental methods. The best solution is to compare the comparable.

This paper contributes to the literature by providing a seemingly unbiased estimate of the effect of cancer on labour supply. While the analysis is restricted to male workers who are least 65 years age, it is important to remember that almost 60% of male cancers per year (484,000 out of 855,000) are diagnosed within this age bracket (American Cancer Society, 2014b). Considering that labour force participation rate for men who are least 65 years age is 24%, up to 120,000 of these workers may be diagnosed per year (Bureau of Labor Statistics, 2013). While the results are not representative of the labour market as a whole, it is important to remember that they are applicable to a non-negligible portion of it.

The rest of the paper is organised as follows. In Section 2, I present medical and epidemiological reasons for risk factor attenuation in my sample. Section 3 contains information on the methodology and the data used in the analysis. Section 4 contains numerous statistical demonstrations of covariate balance between the control and treatment

groups. In Section 5, I present the main results and examine how the effect varies with different subgroups. Section 6 uses propensity score matching to investigate the sensitivity of the results to a potentially omitted confounder. Finally, in Section 7, I assess the mechanism through which the reduction in labour supply takes place by analysing changes in specific levels of health and labour force status. Section 8 then concludes.

2. CANCER AND IT'S RISK FACTORS

Cancer is the name given to a group of diseases who share the common trait of abnormal cell growth which have the potential to cause sickness or death. All cancer is genetic in that it is caused by gene mutation and it usually takes multiple genetic mutations over a lifetime to cause cancer (American Society of Clinical Oncology, 2015b). In about 5 to 10% of cancers, a specific gene mutation is inherited from the parent, while 90 to 95% of mutations are caused by environmental factors. However, the term "environmental" is a misnomer in the sense that it is used to refer to anything that is not hereditary. Some environmental risk factors such as ageing or background radiation are unavoidable. However, some environmental risk factors are avoidable, and these include tobacco, alcohol, poor diet etc.

With this in mind, how much of cancer diagnosis can be attributed to these avoidable risk factors? Regarding the most prominent risk factors for cancer, a 2005 study published in The Lancet shows that, in high income countries, approximately 37% of all cancer deaths are attributable to modifiable risk factors (Danaei et al., 2005). A breakdown of the five largest risk factors is presented in Table 1. Smoking is by far the largest modifiable risk factor, accounting for 29%, while the four other major risk factors (alcohol use, overweight and obesity, low fruit and vegetable intake, and physical inactivity) modestly account for between 2 to 4% each.

Table 1. Behavioural and Environmental Cancer Risk Factors

	Number of Deaths	%
Total number of cancer deaths in high income countries	2,066,388	100%
Total percentage attributable to risk factors	764,564	37%
Total percentage attributable to specific risk factors		
Smoking	596,000	29%
Alcohol use	88,000	4%
Overweight and obesity	69,000	3%
Low fruit and vegetable intake	64,000	3%
Physical inactivity	51,000	2%

Note: Adapted from Danaei et al. (2005)

The idea that age is the biggest risk factor in developing cancer came to prominence over 50 years ago. Both Nordling (1953) and Armitage and Doll (1954) demonstrated an exponential relationship between age and the incidence of cancer. The theory, according to Nordling (1953), is that if a large enough number of cells is allowed a sufficiently long period of time, gene mutations will necessarily occur in some of them. If a cell collects enough mutations it may become cancerous. This is evident today in the age profile of cancer incidence. The graph in Figure 1, constructed with data from the Surveillance, Epidemiology and End Results (SEER) Program, shows the rates of male cancer incidence in the U.S between 1975 and 2011. The average incidence rates per 100,000 for the below age 50, ages 50 to 64, and age 65 and over, are 76, 873 and 2,885 respectively, showing the large increase in incidence through the different age brackets (SEER, 2015).

All of this information is used to construct a sample for which the effect of modifiable risk factors on cancer is restricted. By using a sample of working males, who are at least 65 years of age, and who have never been diagnosed with cancer before, it allows for two things. It attenuates much of the effect of modifiable risk factors on cancer diagnosis simply because of the high risk nature of the group (Rothman and Poole, 1988). This has been demonstrated with relation to both cancer and cardiovascular disease (Odden et al., 2014; Rottenberg et al.,

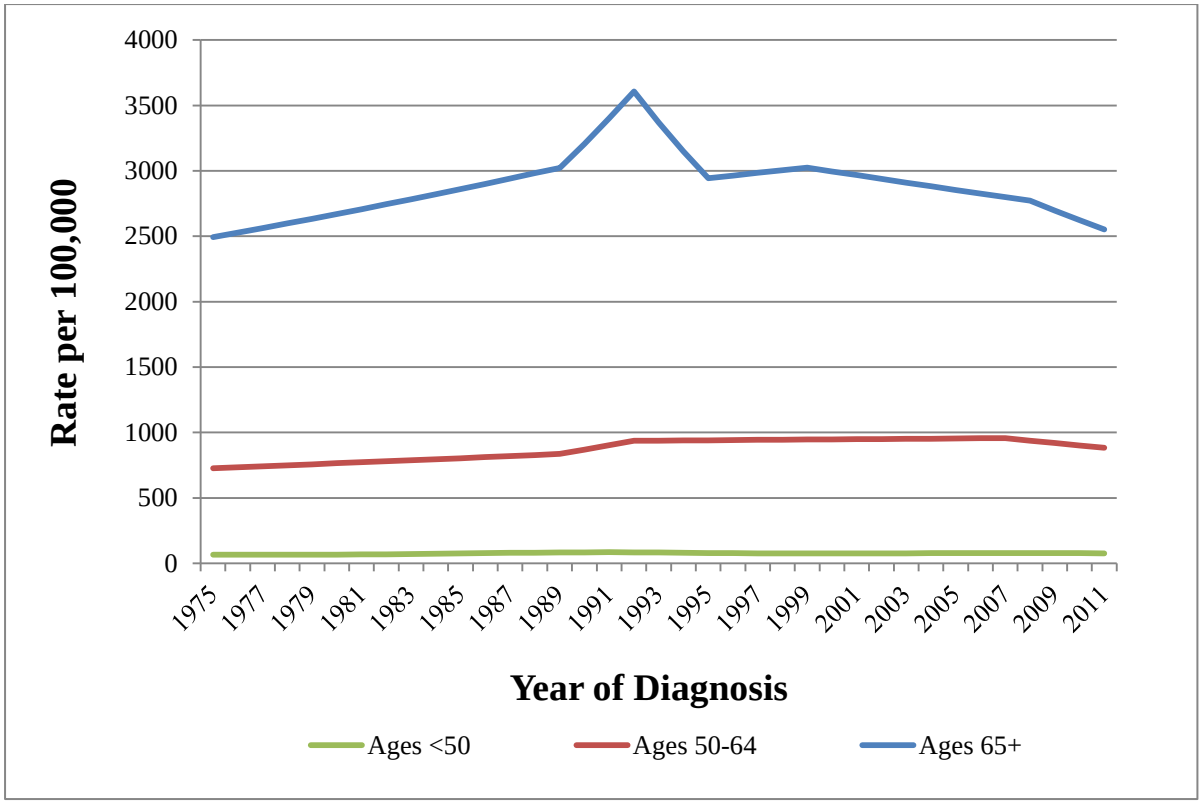

Figure 1. Age-Adjusted Incidence Rates by Age at Diagnosis/Death, All Sites, All Races, Male, 1975 - 2011
Note: Adapted from SEER (2015)

2011; Sweeney et al., 2004). It also results in a 'depletion of the susceptible' because the people who remain in the sample cannot have been diagnosed with cancer before, they cannot have died from any other competing diseases, and they must be healthy enough to remain working.

The advantage of using this approach is that it gives rise to in a group of workers for whom traditional modifiable risk factors have a negligible impact. This means that the effect of cancer on employment varies little with the addition of covariates or the removal of important subgroups, strengthening the view that it is an unbiased effect. However, this approach also has disadvantages, the main one being that the group are a highly selective cohort. Generalizing results from this group to that of all cancer survivors would be inappropriate, and this diminishes the applicability of any policy implications arising from this study.

Essentially, this approach improves the internal validity of the estimates, but it does so at the expense of some external validity.

3. METHODOLOGY AND DATA

3.1 Methodology

Following notation from Angrist and Pischke (2009), cancer diagnosis can be thought of as binary variable, $D_i = \{0,1\}$, and the outcome of interest, employment, is denoted by Y_i. The observed outcome for an individual may be written as a combination of their potential outcomes

$$Y_i = \begin{cases} Y_{1i}, & \text{if } D_i = 1 \\ Y_{0i}, & \text{if } D_i = 0 \end{cases}$$

$$= Y_{0i} + (Y_{1i} - Y_{0i})D_i, \quad (1)$$

where the causal effect of cancer on employment is $Y_{1i} - Y_{0i}$. In a regression framework, this could be represented as

$$Y_i = \alpha + \rho D_i + \eta_i, \quad (2)$$

with $\rho = Y_{1i} - Y_{0i}$. Evaluating the conditional expectation of this equation gives

$$E[Y_i|D_i = 1] = \alpha + \rho + E[\eta_i|D_i = 1] \quad (3)$$

$$E[Y_i|D_i = 0] = \alpha + E[\eta_i|D_i = 0], \quad (4)$$

so that

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \rho + E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0]. \quad (5)$$

In this case, selection bias arises when the expected value of the error term, η_i, is different for the respondents who get cancer and the respondents who do not get cancer and the cov(D_i η_i) is not zero. If the expected values of the error term differ, then control variables, X_i, can be included. Selection bias will be eliminated if

$$E[\eta_i|X_i, D_i = 1] - E[\eta_i|X_i, D_i = 0] = E[\eta_i|X_i, D_i] = 0. \quad (6)$$

Even in cases where cov(D_i η_i) is zero, the control variables should still be included if they explain variation in Y_i and, therefore, reduce residual variance.

With longitudinal data, it is possible to measure the outcome variable as the change in the outcome over time, ΔY_i. In this case,

$$\rho = \Delta Y_{1i} - \Delta Y_{0i}. \quad (7)$$

and ρ can be thought of as a difference-in-differences estimator. This is because it calculates the difference in the outcome differences between the cancer and non-cancer group. This removes any temporally-persistent separable components of bias. For this is analysis, the employment outcomes that I am interested in are the binary measure of whether the respondent is working and the continuous measure of their hours of work

3.2 Data

In order to obtain the requisite data on older workers with which to do this analysis, I use data from the Health and Retirement Study (HRS). The HRS is a large, longitudinal data set which contains information on the respondents' health, wealth, employment and other demographic information. The first wave was collected in 1992 using a nationally representative sample of 51 to 61 year olds and, since then, it has been collected every two years. For this analysis, I use the first 10 waves of data, spanning the years 1992 to 2010.

As with any study which examines the impact that cancer has on the labour force, it is important to observe the respondents' pre-cancer behaviour. To do this, information on the respondents, both before and after they are diagnosed with cancer, is required. This means that information in two different time periods is required for each observation. Having 10 waves of data allows me to observe 9 potential non-cancer to cancer transitions: wave 1 to

wave 2, wave 2 to wave 3, . . . , wave 9 to wave 10. I follow Finkelstein (2004) and 'stack' these 9 time periods into two simple before and after periods, hereafter referred to as Period 1 and Period 2. Combining these periods in the HRS gives a total of 146,173 observations. Table 2 provides more information on how I restrict the sample. Imposing these restrictions leaves a final sample of 3,923 respondents, of which, 180 will be diagnosed with cancer in Period 2. Because this is a stacked panel, individuals may appear more than once. Of the 3,923 observations, 1,819 are unique individuals.

In addition to using the employment and cancer information, I also use information on pre-diagnosis variables that are measured in Period 1. This includes demographic information such as age, age squared, whether the respondent is non-white, has some or a full college education, has more than two people living in the household. The cancer risk factors that are taken into account are whether the respondent is in poor health[1] (since there are no measures of fruit or vegetable intake in the HRS, I use poor health as a proxy 'risk factor'), is a currently a smoker, ever drinks any drinks alcohol, is obese and if they are not getting regular exercise[2]. Regarding their employment situation, I include variables indicating whether they have a working spouse, are self-employed, hours of work per week and whether they have a defined benefit or defined contribution pension. Their earnings tertile and household income quartile are also considered[3]. All the variables are binary except for age, age squared and hours of work. Finally, I include wave identifier and census division dummy variables. Because an individual can appear in more than one observation, the standard errors in the regression analysis are clustered at the individual level.

[1] Poor health is defined as being in poor or fair health as opposed to good, very good or excellent in a self-reported health measure.
[2] Regular exercise is defined as whether you engage in vigorous physical activity at least 4 times a week for waves 1 – 6 or every day for waves 7 – 9.
[3] The earnings and income measures are adjusted to 2009 dollars using the Bureau of Labor Statistics inflation calculator.

Table 2. Sample Information

Exclusion criteria	**Observations**
Unrestricted sample	146,173
Cancer in Period 1[a]	16,888
Not working in Period 1[b]	71,991
Aged below 65 or above 78 in Period 1[c]	46,269
Not married in Period 1[d]	3,348
Female [e]	2,632
Works more than 80 hours a week in Period 1 or Period 2 [f]	427
Contradictory or anomalous information [g]	264
Missing or incomplete survey records regarding control or outcome variables [h]	431
Restricted sample	3,923

Note: [a] I remove respondents who report cancer in Period 1 as we cannot observe their pre-cancer behaviour
[b] This means that the effect that I estimate will be for respondents who are working
[c] I remove those below age 65 because of the identification strategy and those above 78 because of the small number of observations for this group
[d] Because married workers and single workers face different time, income and support constraints, I focus my analysis on married respondents
[e] I also exclude females since there are not enough cancer cases to perform extensive analyses
[f] This reduces the impact of extreme observations on the estimates
[g] This includes reporting working but missing values for the hours of work, 7 or more people in the household, BMI greater than 60, not working for 40 non-consecutive quarters, respondent's whose Period 2 hours of work is more than quadruple their Period 1 hours of work, and those who live outside the U.S. or in a U.S. territory.
[h] This includes answers such as refusal, don't know and other non-coded responses.

4. COVARIATE BALANCE BETWEEN GROUPS

The descriptive statistics for both the cancer and non-cancer groups are presented in Table 3, along with *t*-tests for the equality of their means. Of 30 pre-diagnosis Period 1 variables that are tested, only age and age squared report statistically significant differences at the 10% level. In terms of cancer risk factors, the groups appear balanced. The cancer group are about 5 percentage points less likely to drink alcohol and about 5 percentage points more likely to no have regular exercise, though neither result reaches statistical significance. Looking at the Period 2 outcomes, there is a 13.3 percentage point reduction in employment cancer group, which corresponds to 4.5 fewer hours a week being worked. Both effects are significant at the 1% level.

Table 3. Descriptive Statistics

Variable	Cancer Group	Non-Cancer Group
Period 1		
Demographics		
Age	*69.71	69.17
Age squared	*4871	4797
Non-white	11.67%	10.29%
College	49.44%	47.18%
More than two people living at home	26.67%	21.67%
Cancer Risk Factors		
Poor health	16.11%	14.43%
Smoker	10.56%	10.55%
Alcohol drinker	54.44%	59.79%
Obese	24.44%	24.15%
No regular exercise	69.44%	67.19%
Employment		
Spouse working	42.78%	42.64%
Self-employed	38.89%	39.25%
Hours of work (per week)	30.11	30.48
Defined benefit pension	9.44%	7.53%
Defined contribution pension	11.67%	13.89%
Earnings Tertiles (1st Omitted)		
2nd Earnings tertile	34.44%	32.67%
3rd Earnings tertile	30.56%	33.69%
Household Income Quartiles (1st Omitted)		
2nd Household income quartile	30.00%	24.69%
3rd Household income quartile	25.00%	24.93%
4th Household income quartile	24.44%	25.25%
Wave Number (1, 2, or 3 Omitted)		
Wave 4, 5, or 6	40.56%	41.44%
Wave 7, 8, or 9	45.56%	43.73%
Census District (New England Omitted)		
Mid-Atlantic	10.56%	10.29%
East North Central	17.22%	17.29%
West North Central	8.89%	11.84%
South Atlantic	25.56%	24.02%
East South Central	6.11%	5.64%
West South Central	9.44%	10.55%
Mountain	6.11%	5.18%
Pacific	11.11%	11.22%
Period 2		
Employment		
Working	***60.00%	73.31%
Hours of work (per week)	***16.83	21.40
Observations	180	3,743

Note:
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

Another way of testing for covariate balance is to follow Bailey et al. (2012) and regress the pre-diagnosis covariates on future cancer diagnosis. The additional benefit of using this test is that the wave and district dummy variables from Table 3 can be included to reduce any residual variance. Thus, the coefficient on cancer measures the residual correlation between cancer and pre-diagnosis characteristics that could indicate correlations with other, unobserved characteristics. This provides a more rigorous test of covariate balance. The results of each regression are presented in Table 4. Again, the only two models which give statistically significant results are age and age squared, though in this case, the results are significant at the 5% level. However, even with the reduction in residual variance that the extra dummies provide, cancer is not significantly related to any of the pre-diagnosis variables, including the traditional risk factors.

In order to test the joint significance of the covariates with relation to cancer, I regress future cancer diagnosis on all the other covariates in the model in Table 5. As expected, because of the lack of correlation between the other variables and cancer diagnosis, an *F*-test for the joint significance of the variables cannot be rejected, meaning that there is no evidence that they help explain future cancer diagnosis. Given the lack of a relationship between cancer diagnosis and its known risk factors, it may be tempting to say that the variables used in the analysis are poor measures of the true variables. I attempt to address this issue in the remaining columns in Table 5. Here, I regress future diagnosis of high blood pressure, heart problems, lung disease, arthritis and diabetes on the same variables. It is apparent that, even after controlling for a previous diagnosis, four of the five known risk factors (except for whether the respondents drink alcohol) are statistically significant in determining these diseases. Since these variables help explain variation in future diagnoses of other diseases,

Table 4. The Effect of Cancer on Pre-Treatment Characteristics

	Age	*Age squared*	*Non-white*	*College*	*More than two people living at home*
Cancer	0.538**	74.442**	0.012	0.020	0.049
	(0.264)	(37.335)	(0.024)	(0.037)	(0.033)
Observations	3,923	3,923	3,923	3,923	3,923
R-squared	0.021	0.021	0.012	0.023	0.026
	Poor health	*Smoker*	*Alcohol drinker*	*Obese*	*No regular exercise*
Cancer	0.018	0.001	-0.057	0.001	0.013
	(0.028)	(0.023)	(0.037)	(0.032)	(0.029)
Observations	3,923	3,923	3,923	3,923	3,923
R-squared	0.011	0.013	0.035	0.023	0.304
	Spouse working	*Self-employed*	*Hours of work (per week)*	*Defined benefit pension*	*Defined contribution pension*
Cancer	0.001	-0.001	-0.310	0.019	-0.024
	(0.037)	(0.037)	(1.207)	(0.022)	(0.025)
Observations	3,923	3,923	3,923	3,923	3,923
R-squared	0.005	0.010	0.013	0.012	0.009
	2^{nd} Earnings tertile	*3^{rd} Earnings tertile*	*2^{nd} Household income quartile*	*3^{rd} Household income quartile*	*4^{th} Household income quartile*
Cancer	0.019	-0.035	0.055	0.000	-0.011
	(0.036)	(0.035)	(0.034)	(0.033)	(0.033)
Observations	3,923	3,923	3,923	3,923	3,923
R-squared	0.005	0.014	0.007	0.005	0.011

Note: All models estimated by OLS regression.
All models also include the wave and district dummy variables from Table 3.
Clustered standard errors (by person) in parentheses.
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

it supports the idea that the covariates are in fact balanced between the cancer and non-cancer groups, rather than the idea that the variables are simply coarse measures.

Table 5. Selection into Cancer versus Selection into Other Diseases

Variables **(Cancer Risk Factors in bold)**	Cancer	High Blood Pressure	Heart Problems	Lung Disease	Arthritis	Diabetes
Reports condition in Period 1	-	0.904***	0.925***	0.972***	0.896***	0.952***
	-	(0.007)	(0.006)	(0.004)	(0.007)	(0.005)
Age	0.049	0.015	-0.013	-0.004	-0.059	0.062**
	(0.036)	(0.038)	(0.041)	(0.023)	(0.039)	(0.029)
Age^2	-0.000	-0.000	0.000	0.000	0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-white	0.004	-0.008	-0.029***	-0.008	-0.000	0.012
	(0.012)	(0.010)	(0.010)	(0.007)	(0.013)	(0.010)
College	0.006	-0.009	0.009	0.001	-0.007	0.004
	(0.007)	(0.007)	(0.007)	(0.005)	(0.008)	(0.006)
More than two people living at home	0.014	-0.005	0.003	0.003	-0.001	-0.006
	(0.009)	(0.008)	(0.009)	(0.006)	(0.008)	(0.006)
Poor health	**0.006**	**0.012**	**0.044*** **	**0.013**	**0.025** **	**-0.004**
	(0.010)	**(0.009)**	**(0.012)**	**(0.008)**	**(0.010)**	**(0.008)**
Smoker	**0.003**	**0.009**	**-0.013**	**0.042*** **	**-0.001**	**-0.003**
	(0.011)	**(0.012)**	**(0.011)**	**(0.012)**	**(0.011)**	**(0.009)**
Alcohol drinker	**-0.010**	**0.001**	**-0.010**	**0.002**	**-0.006**	**-0.008**
	(0.007)	**(0.007)**	**(0.007)**	**(0.005)**	**(0.007)**	**(0.006)**
Obese	**0.001**	**0.015***	**0.007**	**0.009**	**0.019** **	**0.036*** **
	(0.008)	**(0.008)**	**(0.008)**	**(0.006)**	**(0.009)**	**(0.008)**
No regular exercise	**0.004**	**0.025*** **	**0.014**	**0.000**	**-0.002**	**0.019*** **
	(0.008)	**(0.008)**	**(0.009)**	**(0.005)**	**(0.009)**	**(0.007)**
Spouse working	-0.000	-0.017**	0.005	0.011**	-0.004	-0.005
	(0.007)	(0.007)	(0.007)	(0.004)	(0.007)	(0.005)
Self-employed	-0.004	0.000	-0.003	-0.005	-0.005	-0.003
	(0.009)	(0.008)	(0.009)	(0.007)	(0.010)	(0.007)
Hours of work (per week)	0.000	0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Defined benefit pension	0.013	-0.013	-0.018	-0.003	-0.009	0.004
	(0.014)	(0.011)	(0.012)	(0.007)	(0.013)	(0.011)
Defined contribution pension	-0.004	0.013	-0.010	0.001	0.010	-0.001
	(0.010)	(0.011)	(0.010)	(0.006)	(0.011)	(0.008)
2^{nd} Earnings tertile	-0.003	0.010	0.005	-0.018**	0.013	-0.008
	(0.011)	(0.009)	(0.010)	(0.007)	(0.011)	(0.008)
3^{rd} Earnings tertile	-0.012	0.001	0.006	-0.007	-0.018	-0.009
	(0.011)	(0.010)	(0.010)	(0.008)	(0.011)	(0.008)
2^{nd} Household income quartile	0.021**	0.015	-0.008	-0.006	0.002	-0.013
	(0.010)	(0.010)	(0.010)	(0.007)	(0.010)	(0.008)
3^{rd} Household income quartile	0.015	0.012	-0.010	-0.011	0.019*	-0.009
	(0.011)	(0.010)	(0.010)	(0.007)	(0.011)	(0.009)
4^{th} Household income quartile	0.015	0.011	-0.005	-0.005	0.013	-0.023**
	(0.012)	(0.011)	(0.012)	0.972***	(0.012)	(0.009)
Observations	3,923	3,917	3,922	3,922	3,918	3,918
F-statistic	0.74	892	1,657	7,689	757	4,146
R-squared	0.006	0.832	0.794	0.744	0.820	0.828

Note: All models estimated by OLS regression.
All models also include the wave and district dummy variables from Table 3.
Clustered standard errors (by person) in parentheses.
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

As a final test of balance between the two groups, I plot how the health status of both groups changes over time. One of the benefits from using the HRS data is that, for a subsample of respondents, I get to observe their health status in the two years before Period 1. I refer to this new period as Period 0. For 3,649 respondents (173 of which will get cancer in Period 2) information for Period 0, Period 1 and Period 2 is available and I use this information to plot the proportion of the sample that reports being in poor health. This means that for the 3,476 non-cancer respondents, I now have information on health status for three periods, during which none of them are diagnosed with cancer. For the 173 cancer respondents, I now have information on their health status for two pre-diagnosis periods and one post-diagnosis period.

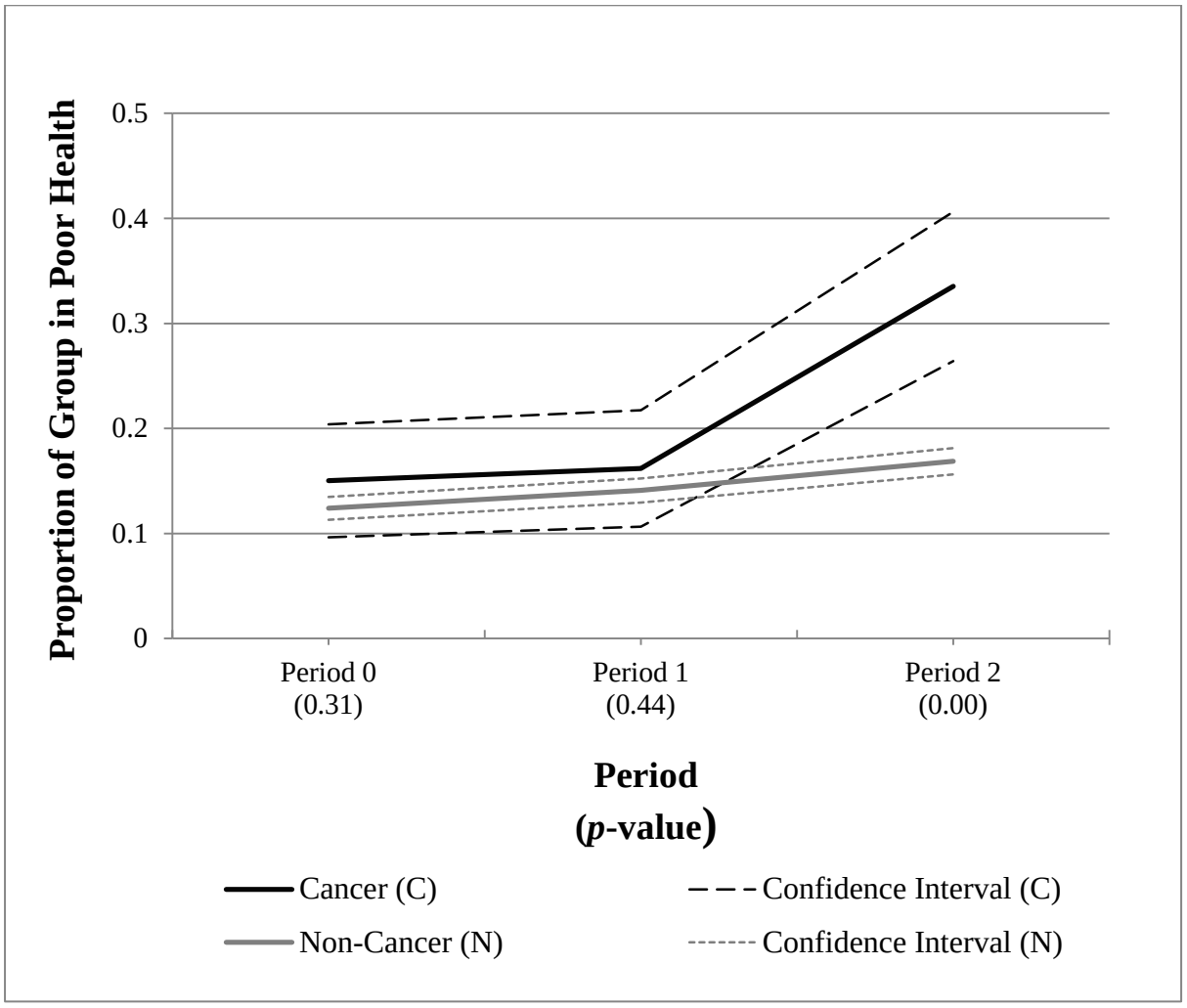

Figure 2. Changes in the Proportion of Respondents in Poor Health over Time

In Figure 2, the proportion of each group in poor health is plotted against the three time periods. In Period 0 and Period 1, a higher proportion of the cancer group are in poor health, though neither difference is meaningfully or statistically significant (15.0% vs 12.4% and 16.2% vs. 14.1% respectively). In Period 2, after the respondents have been diagnosed with cancer, the proportion of respondent in poor health in the cancer group is double that of the non-cancer group (33.5% vs. 16.9%, $p<0.01$). This lends more credence to the notion that the groups are comparable and any difference arising in Period 2 is due to cancer diagnosis.

5. RESULTS

5.1 Main Results

The results from the main analysis are presented in Table 6. The first column contains the results from a simple unadjusted OLS regression of whether the respondent is working in Period 2 on whether they have been diagnosed with cancer. The cancer group are 13.3 percentage points ($p<0.01$) less likely to be working than the non-cancer group. The second column adds in the pre-diagnosis variables to the regression. In this case, the cancer group are 13.2 percentage points ($p<0.01$) less likely to be working than the non-cancer group.

In the third and fourth columns, I use the change in hours of work per week as another measure of employment outcomes. This estimate is not conditional on the respondents returning to work in Period 2 and so captures the reduction in hours at the extensive margin. In order to avoid constructing a mechanical endogeneity problem, the Period 1 measure of hours of work is excluded as a control variable in these models. The unadjusted model shows that respondents diagnosed with cancer work 4.2 ($p<0.01$) fewer hours per week than the non-cancer group. This rises to 4.4 ($p<0.01$) fewer hours per week in the adjusted model.

Table 6. The Effect of Cancer on Employment: Main Results

Variable	Working (Unadjusted)	Working (Adjusted)	Hours of Work (Unadjusted)	Hours of Work (Adjusted)
Cancer	-0.133***	-0.132***	-4.203***	-4.365***
	(0.037)	(0.036)	(1.313)	(1.282)
Age		0.067		-0.314
		(0.077)		(2.864)
Age^2		-0.001		0.003
		(0.001)		(0.020)
Non-white		-0.019		0.262
		(0.025)		(0.881)
College (some or full)		0.042**		1.613***
		(0.017)		(0.586)
More than two people living at home		0.038**		0.930
		(0.018)		(0.696)
Poor health		-0.060***		-2.393***
		(0.022)		(0.811)
Smoker		-0.061**		-2.218**
		(0.027)		(1.004)
Alcohol drinker		0.005		0.348
		(0.016)		(0.587)
Obese		-0.022		-1.152*
		(0.017)		(0.643)
No regular exercise		-0.010		-0.173
		(0.018)		(0.694)
Spouse working		0.007		0.419
		(0.015)		(0.570)
Self-employed		0.102***		2.422***
		(0.019)		(0.736)
Hours of work (per week)		0.003***		-
		(0.001)		-
Defined benefit pension		-0.046		-3.389***
		(0.028)		(1.151)
Defined contribution pension		0.019		-1.588*
		(0.022)		(0.832)
2nd Earnings tertile		0.093***		3.299***
		(0.022)		(0.834)
3rd Earnings tertile		0.054**		0.003
		(0.022)		(0.873)
2nd Household income quartile		0.034		1.001
		(0.022)		(0.778)
3rd Household income quartile		0.045**		0.139
		(0.022)		(0.813)
4th Household income quartile		0.086***		0.897
		(0.024)		(0.951)
Observations	3,923	3,923	3,923	3,923
R-squared	0.004	0.049	0.003	0.030

Note: All models estimated by OLS regression.
Adjusted models also include the wave and district dummy variables from Table 3.
Period 1 hours of work not included as a control in the Hours of Work model
Clustered standard errors (by person) in parentheses.
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

Since the hours of work model just serves to mirror the working model (with a different measure) I will only refer to the results from the working models from now on.

The fact that the addition of the pre-diagnosis variables results in the adjusted models results in a negligible change to estimates is to be expected, given their lack of correlation with cancer diagnosis. What I do in the remaining subsections, is to see how these estimates vary by removing important subgroups. By checking to make sure that the results are not driven by one aberrant subgroup, it reduces any the possibility of any threat to the validity to the estimates. I first re-estimate the models in Table 6 while removing each risk factor at a time. Also, because cancer may have different effects depending on the respondent's labour market situation, I re-estimate the results while focusing of different working groups.

5.2 Risk Groups

In Table 7, I exclude respondents who report being in poor health in Period 1. Even though the estimate is still statistically significant at the 1%, the coefficient drops 2 percentage points in size. This is not unexpected, as we would assume cancer has less of an effect on healthier workers. Despite the large variation in sample size, excluding smokers, obese respondents, and respondents who drink alcohol has very little effect on the estimates, though the results where the alcohol drinkers are removed are only significant at the 5% level in the working models. In the final specification, respondents who do not exercise regularly are excluded. This restriction results in the largest change to the estimates, as the cancer group are now only 5.7 percentage points less likely to be working than the non-cancer group and none of the estimates are statistically significant. However, any meaningful inference that can be drawn from this group is limited as the sample used in the analysis represents less than a third of the original sample. Also the measure of regular exercise, discussed in Section 3,

Table 7. The Effect of Cancer on Employment: Risk Groups

Variable	Working (Unadjusted)	Working (Adjusted)	Hours of Work (Unadjusted)	Hours of Work (Adjusted)
	Excluding Respondents in Poor Health in Period 1			
Cancer	-0.109***	-0.113***	-2.971**	-3.216**
(Observations = 3,354)	(0.040)	(0.039)	(1.365)	(1.330)
	Excluding Respondents who are Smokers in Period 1			
Cancer	-0.139***	-0.133***	-4.026***	-4.120***
(Observations = 3,509)	(0.039)	(0.038)	(1.354)	(1.335)
	Excluding Respondents who are Obese in Period 1			
Cancer	-0.136***	-0.134***	-3.580**	-3.595**
(Observations = 2,975)	(0.043)	(0.042)	(1.486)	(1.471)
	Excluding Respondents who Drink Alcohol in Period 1			
Cancer	-0.131**	-0.126**	-4.176*	-4.134*
(Observations = 1,587)	(0.056)	(0.054)	(2.143)	(2.120)
	Excluding Respondents who do not Exercise Regularly in Period 1			
Cancer	-0.052	-0.057	0.031	-0.218
(Observations = 1,283)	(0.063)	(0.062)	(2.442)	(2.409)

Note: All models estimated by OLS regression.
Adjusted models also include the control variables from Table 3.
Clustered standard errors (by person) in parentheses.
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

is extremely severe, and may not be applicable to this age group. Nevertheless, the results are included for the sake of completeness.

5.3 Employment Groups

It is quite possible that cancer will have a different effect on respondents depending on their attachment to the labour market. In this analysis, I focus on workers who have a strong attachment to the labour market and, theoretically, should be affect the least by cancer diagnosis. The results are available in Table 8. For example, if we just consider respondents who are working full-time or are self-employed, respondents diagnosed with cancer are 9.8

Table 8. The Effect of Cancer on Employment: Employment Groups

Variable	Working (Unadjusted)	Working (Adjusted)	Hours of Work (Unadjusted)	Hours of Work (Adjusted)
	Respondents who Work Full-Time in Period 1			
Cancer	-0.104*	-0.098*	-5.744**	-5.631**
(Observations = 1,772)	(0.054)	(0.052)	(2.312)	(2.262)
	Respondents who are Self-Employed in Period 1			
Cancer	-0.104*	-0.117**	-4.574**	-5.079**
(Observations = 1,539)	(0.057)	(0.055)	(1.985)	(2.006)
	Respondents who have a Working Spouse in Period 1			
Cancer	-0.104*	-0.103*	-2.914	-3.168
(Observations = 1,673)	(0.055)	(0.056)	(1.997)	(1.965)
	Excluding Respondents who have a College Education			
Cancer	-0.105**	-0.104**	-2.284	-2.448
(Observations = 2,068)	(0.053)	(0.052)	(1.980)	(1.938)

Note: All models estimated by OLS regression.
Adjusted models also include the control variables from Table 3.
Clustered standard errors (by person) in parentheses.
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

percentage points and 11.7 percentage points less likely to work than non-cancer respondents. When focusing exclusively of respondents who have working spouses, which may give them a higher attachment to the labour market based on assortative mating, cancer respondents are 10.3 percentage points less likely to work than non-cancer respondents. As expected, the size of the coefficients are smaller than in the main results, yet despite the strong attachment that these respondents have to the labour market, the estimates are still significant at the 10% level. Also, respondents with lower levels of education are likely to be more attached to the labour force as they have poorer paid jobs, lower Social Security payments if they do retire, or fewer safety nets in the event of cancer diagnosis and, thus, are required to work after diagnosis. Again, when focusing exclusively on this group, the effect falls to 10.4 percentage points.

6. PROPENSITY SCORE MATCHING

6.1 Kernel Matching

As a final robustness test, I estimate the effect of cancer on employment using propensity score matching (PSM). Because PSM relies on the same assumptions as OLS regression, I expect these estimates to be similar to the previous ones. However, one of the benefits of using PSM is that it allows for experimentation with different matching specifications and excluding different percentiles of the propensity score. Because my sample has many non-cancer observations per cancer observation, I use kernel matching, which has been shown to work well in these circumstances (Frölich, 2004). To do this, I use the ATTK commands developed by Becker and Ichino for Stata (Becker and Ichino, 2002). The standard errors are then bootstrapped since bootstrapping kernel matching estimates is still assumed to provide asymptotically valid inference, unlike nearest neighbour matching (Abadie and Imbens, 2008). To select the bandwidth, I follow Heckman et al. (1998) and use a rule of thumb. In this case, the rule of thumb is defined as

$$bandwidth \approx 1.06\, S\, n^{-\frac{1}{5}}, \tag{8}$$

where n is the number of observations in the control group and S is the sample standard deviation of the control group propensity score (Sheather, 2004).

The results for the kernel matching are presented in Table 9. The rule of thumb above suggests a bandwidth of 0.0032[4]. To assess the sensitivity of the results, I estimate the effect using half and double this bandwidth, and also exclude the top and bottom percentiles and deciles. In most cases, the doubling of the bandwidth has little effect on the estimates. Halving the bandwidth does increase the negative effect from cancer diagnosis in nearly all

[4] This rule of thumb provides the optimal bandwidth for a Gaussian kernel. In this case, I am using the Epanechnikov kernel. While this should rule still provide a close to optimal bandwidth for this kernel, I vary the bandwidth to examine its sensitivity.

Table 9. The Effect of Cancer on Employment: Kernel Matching

		Employment	
Propensity Score Restrictions	Bandwidth 0.0016	Bandwidth 0.0032	Bandwidth 0.0064
None (Observations = 3,923)	-0.138*** (0.032)	-0.134*** (0.032)	-0.134*** (0.032)
Excluding 1st and 100th Percentile (Observations = 3,844)	-0.140*** (0.033)	-0.138*** (0.033)	-0.136*** (0.033)
Excluding 1st and 10th Decile (Observations = 3,138)	-0.145*** (0.036)	-0.144*** (0.036)	-0.144*** (0.036)

Note: All models estimated by propensity score matching.
Bootstrapped standard errors (50 replications) in parentheses.
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

cases, yet they are still only moderate increases. Of more interest however, is the fact that the estimates unambiguously increase in size as the extreme values of the propensity scores are excluded. These restrictions on the propensity score values are what Black and Smith (2004) define as 'thick support'. They suggest that when larger effects are found using this method, that it could be indicative of a simple heterogeneous treatment effect, measurement error with respect to the treatment variable, or lingering selection on observables. Given the small number of treatment observations, the distribution of the propensity score is positively skewed, meaning that the mass of the distribution is concentrated on the left side. Because these restrictions apply to the tail ends of the distribution, the restrictions are more likely to exclude respondents who are very likely to not be diagnosed with cancer (close to zero), rather than respondents who are very likely to be diagnosed with cancer (close to one). If the respondents who are the least likely to get cancer are also the respondents most likely to stay working, then their exclusion would explain the increase in the magnitude of the effect. The issue of whether this can be attributed to selection on unobservables is addressed in the next subsection.

6.2 Unobserved Confounder Test

In this scenario, I hypothesize the presence of an unobserved binary confounding variable, U_i, whose exclusion results in selection bias. In this case, the expected value of the error terms are equivalent if U_i is taken into account but otherwise they differ

$$E[\eta_i|X_i, D_i = 1] \neq E[\eta_i|X_i, D_i = 0] \quad (9)$$

$$E[\eta_i|U_i, X_i, D_i = 1] = E[\eta_i|U_i, X_i, D_i = 0]. \quad (10)$$

This will happen if U_i not only selects respondents into cancer diagnosis, hereafter referred to as the treatment, (*s*), but also leads to worse labour market outcomes for those respondents who are not in the treatment group (*d*). If

$$p_{ij} \equiv Pr(U = 1|\, D = i, Y = j, X) = Pr(U = 1|\, D = i, Y = j), \quad (11)$$

$$p_{i.} \equiv \sum_{j=0}^{1} p_{ij} \cdot \Pr(Y = j|T = i), \quad (12)$$

with $i,j \in \{0, 1\}$, then this can be expressed more precisely as $s = p_{1.} - p_{0.} > 0$ (the probability of having the confounder given you are in the treated group is greater than the probability of having the confounder given you are in the untreated group) and $d = p_{01} - p_{00} < 0$ (the probability of having the confounder given you have a negative work outcome when in the untreated group is greater than the probability of having the confounder given you have a positive work outcome in the untreated group)[5].

In the top panel of Table 10, I hypothesise that the unobserved confounder takes on a distribution similar to some of the binary variables already in the model. For example, if the omitted variable were to take on a distribution similar to the poor health variable then it would result in the estimate of cancer on working falling from -13.4 percentage points to -13.2 percentage points, a minor attenuation. In fact, this is the largest attenuation that happens

[5] See both Ichino et al. (2008) and Nannicini (2007) for a more rigorous explanation of the tests and the SENSATT command to implement them in Stata.

Table 10. Sensitivity of the Working Kernel Matching Estimate to an Unobserved Confounder

Original Matching Estimate = -0.134

'Calibrated' confounder

Confounder like	New Estimate	*d*	*s*
Poor health	-0.132	-0.05	0.02
Smoker	-0.133	-0.05	0.00
Alcohol drinker	-0.133	0.04	-0.06
Obese	-0.134	-0.02	0.00
No regular exercise	-0.133	0.02	-0.01
Non-white	-0.133	-0.02	0.02
College	-0.135	0.10	0.02
More than two people living at home	-0.135	0.02	0.05
Spouse working	-0.134	0.05	0.00
Self-employed	-0.134	0.09	0.00

'Killer' confounder

Confounder like	New Estimate	*d*	*s*
Manually constructed	-0.120	-0.10	0.10
Manually constructed	-0.106	-0.10	0.20
Manually constructed	-0.097	-0.10	0.30
Manually constructed	-0.105	-0.20	0.10
Manually constructed	-0.083	-0.20	0.20
Manually constructed	-0.056	-0.20	0.30
Manually constructed	-0.094	-0.30	0.10
Manually constructed	-0.056	-0.30	0.20
Manually constructed	-0.022	-0.30	0.30

Note:
$d = p_{01} - p_{00}$, The probability of having the confounder given you have a positive work outcome when in the untreated group minus the probability of having the confounder given that you have a negative work outcome when in the untreated group.
$s = p_{1.} - p_{0.}$, The probability of having the confounder given that you are in the treated group minus the probability of having the confounder given that you are in the untreated group.

when the omitted variable takes on a distribution on one of the binary variables already in the model. Again, this is to be expected given weak relationship between these variables and cancer diagnosis.

I also vary the values of both *d* and *s* manually and examine how the estimate changes in response. This allows me to demonstrate how strong both of the effects would have to be in order to drive the estimate to zero. In order to reduce the dimensionality of the problem in the search for the 'killer' confounders, I need a system of four equations to identify the four p_{ij}

parameters that characterise the confounder's distribution. To do this, I fix at pre-determined values the parameters $Pr(U = 1)$ [the prevalence of the confounder in the whole sample] and $p_{11} - p_{10}$ (the difference in the probability of having the confounder given that you have a positive labour market outcome in the presence of the treatment and the probability of having the confounder given that you have a negative labour market outcome in the presence of the treatment). The values they are fixed at are

$$\begin{aligned} Pr(U = 1) = p_{11} \cdot Pr(Y = 1|T = 1) \cdot \Pr(T = 1) + \\ p_{10} \cdot Pr(Y = 0|T = 1) \cdot \Pr(T = 1) + \\ p_{01} \cdot Pr(Y = 1|T = 0) \cdot \Pr(T = 0) + \\ p_{00} \cdot Pr(Y = 0|T = 0) \cdot \Pr(T = 0) = 0.2. \end{aligned} \quad (13)$$

$$p_{11} - p_{10} = 0. \quad (14)$$

These values indicate that the prevalence of the confounder in the sample in 20%, while the effect of the confounder on the treated outcome is normalised to zero. Ichino et al. (2008) note that these parameters are not expected to threaten the validity of the estimates and so I hold them at fixed known values in order to simulate value of *d* and *s*.

In the second panel of Table 10, the values of both *d* and *s* that are required to produce substantial changes in the estimate are far greater than any variable already in the model. This means that there would have to be an unobserved variable which strongly makes the respondents more likely to get cancer and less likely to work if they did not have cancer. Given the major cancer risk factors and other important determinants of work are already included in the model, this seems unlikely. This test provides the most compelling evidence of the absence of selection bias in the estimates.

7. HEALTH VERSUS PREFERENCES

So far, cancer has shown an unambiguously negative effect on employment. But how is this effect manifesting itself? Figure 3 shows the labour force status of both groups in Period 1 and Period 2. While the labour force status of the two groups is balanced in Period 1, a higher proportion of the cancer group are retired in Period 2. This increase in retirement is to be expected, although the reason that these respondents are retiring may be unclear. Has cancer negatively affected their health so much that they cannot work or have their preferences for work been changed by a large health shock? Figure 4 shows the health status of both groups in Period 1 and Period 2. It is evident that a higher proportion of the cancer group report being in poor or fair health after cancer diagnosis. However, respondents may over report being in poor health as a justification for their retirement decision (Anderson and Burkhauser, 1985; Bound, 1991). What we observe may not be a change in health but a change in preferences manifesting as health changes.

In order to explore this issue further, I examine the effect of cancer on employment for respondents who do not report being in poor health in either Period 1 or Period 2. The results are presented in Table 11. In this case, the respondents who are diagnosed with cancer are still less likely to work than the non-cancer respondents, though the estimate is only 7.8 percentage points ($p<0.1$). Considering that this group are the least likely to be affected by cancer diagnosis, it suggests that there is some evidence that cancer diagnosis is causing a change in preferences in addition to any health changes.

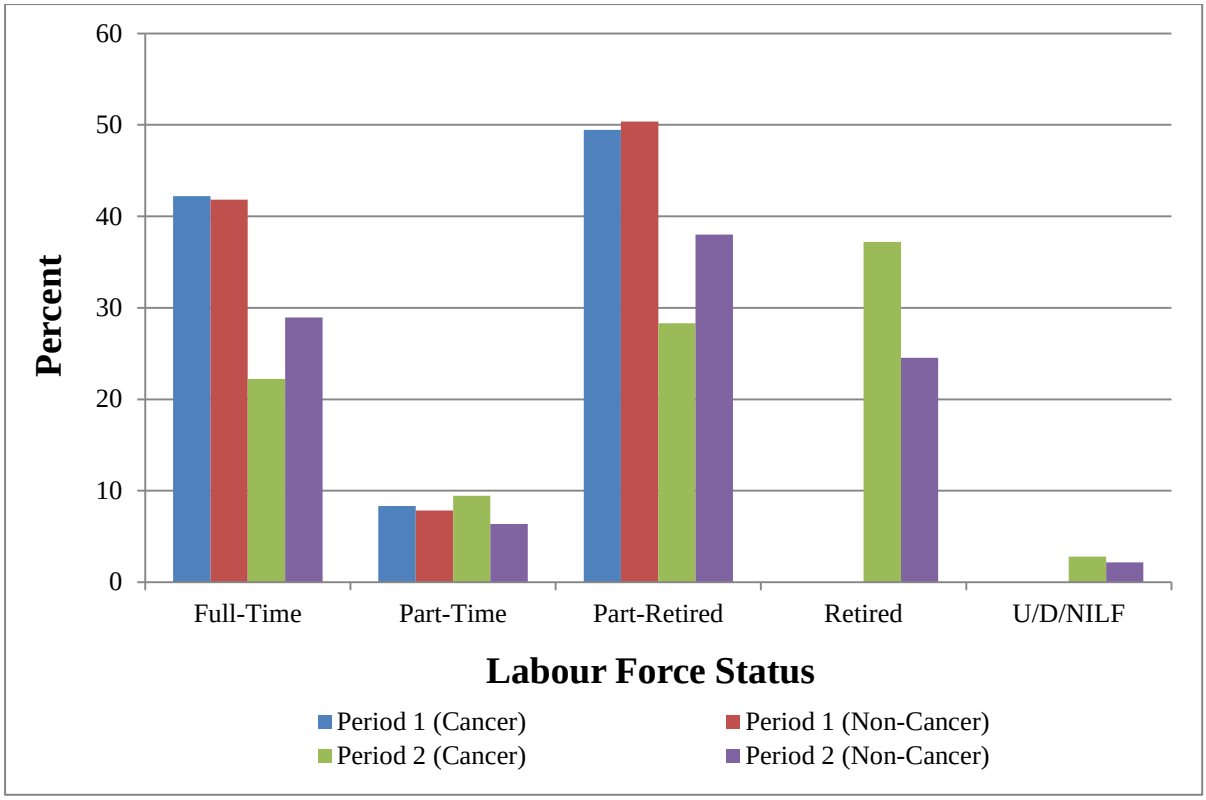

Figure 3. Labour Force Status of Cancer and Non-Cancer Groups
Note: U/D/NILF, Unemployed, Disabled, Not in Labour Force

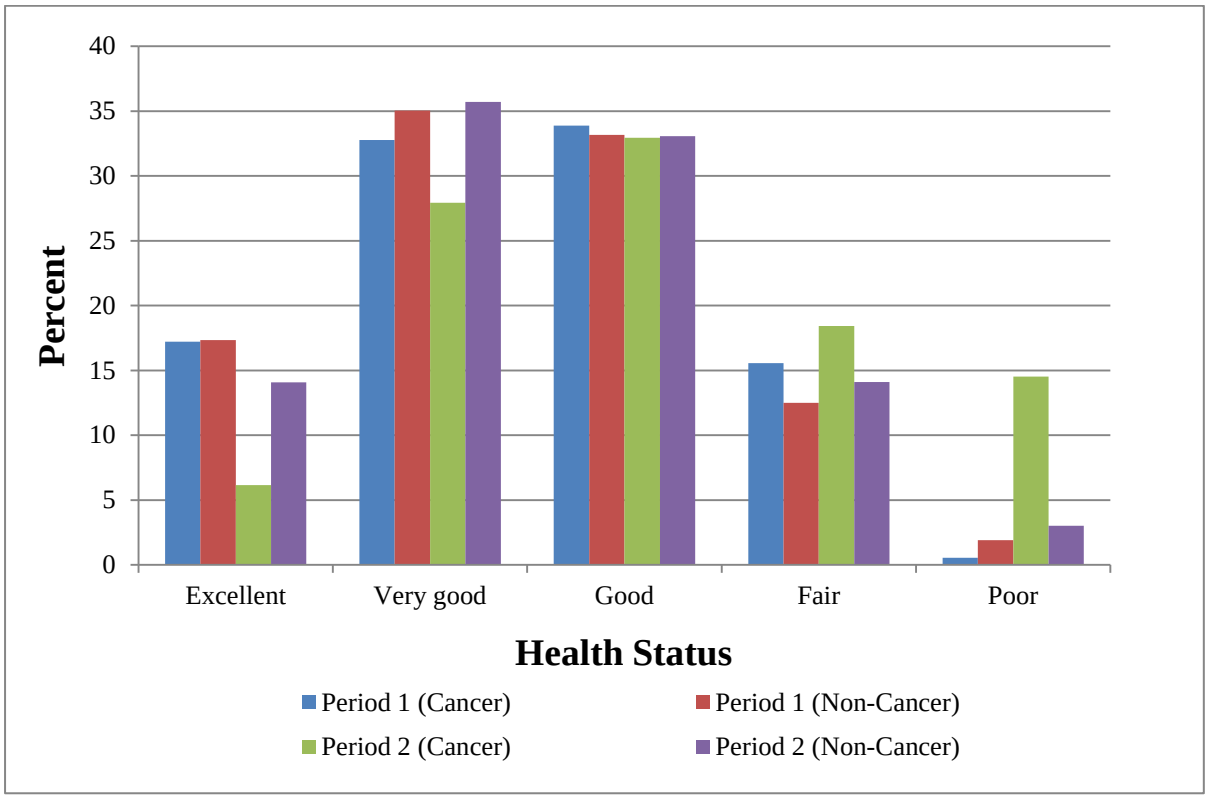

Figure 4. Health Status of Cancer and Non-Cancer Groups

Table 11. The Effect of Cancer on Employment: Healthy Workers

Variable	Working (Unadjusted)	Working (Adjusted)	Hours of Work (Unadjusted)	Hours of Work (Adjusted)
	Respondents who are not in Poor Health in either Period			
Cancer	-0.070	-0.078*	-1.427	-1.717
(Observations = 2,986)	(0.044)	(0.043)	(1.561)	(1.525)

Note: All models estimated by OLS regression.
Adjusted models also include the control variables from Table 3.
Clustered standard errors (by person) in parentheses.
* Result significant at the 10% level.
** Result significant at the 5% level.
*** Result significant at the 1% level.

8. CONCLUSION

Estimating the unbiased effects of health shocks on employment is an important topic in both health and labour economics. This is particularly relevant to cancer, where improvements in screening and treatments have led to increases in survival for nearly all types of cancer. In order to reduce the issue of selection bias, I estimate the effect of cancer on employment for a high-risk cancer group, thus attenuating the effect of many cancer risk factors. While this results in a select sample with which to conduct the analysis (thus, reducing the extremal validity of the results), it does succeed in balancing the observable pre-diagnosis characteristics of both the cancer and non-cancer groups, increasing the likelihood that the unobservable characteristics are also balanced (thus increasing the interval validity of the results).

Respondents who are diagnosed with cancer are 13.2 percentage points less likely to work than their non-cancer counterparts. Both subgroup analysis and PSM suggest that cancer diagnosis has a negative effect on employment, though the effect is bounded between 10 to 14 percentage points. There is also rudimentary evidence that the reduction in employment is

part driven by changes in the respondents' preferences for work and is not just simply a physical health shock.

The question remains of how much external validity is sacrificed in order to obtain these unbiased estimates? Given that almost 60% of male cancers per year are diagnosed in men over the age of 65, this age range is certainly representative of people diagnosed with cancer. However, the labour force participation rate for men who are between the ages of 65 and 74 is only 32%, compared to 70% for those between the ages of 55 to 64 (Bureau of Labor Statistics, 2013). Is this due to certain types of workers selecting into retirement? While workers who work past the age of benefit availability tend to be healthier, better educated, and work in less stressful jobs, there is substantial overlap in these characteristics between workers and retirees (Aaron and Callan, 2011). This suggests that some of the decision can be attributed to preferences or tastes for work.

In light of this, these results have meaningful implications for policy. The normal retirement age will be raised to 67 in 2022. There is a possibility that it could be raised to 70 by 2040 (Wittenburg et al., 2000). This means that in the future, more people who are at least 65 years of age will be diagnosed with cancer while working. If there is little difference between the people who work past the age of benefit availability and the people who retire, then these results could provide important information on how future workers (who normally would have retired) will behave when diagnosed with cancer while in employment. Also, with the normal retirement age increasing, there is evidence to suggest that workers are switching to Social Security Disability Insurance (SSDI) to compensate for potential lost income (Duggan et al., 2007). Considering that cancer diagnosis causes a 13.2 percentage point decrease in the probability of employment over and above people who are retiring anyway, policy makers

should be aware that the number of SSDI claimants due to cancer diagnosis is likely to increase in the future.

REFERENCES

Aaron, H.J., Callan, J.M., 2011. Who retires early?, Working Paper 2011-10. Boston College.

Abadie, A., Imbens, G., 2008. On the Failure of the Bootstrap for Matching Estimators. Econometrica 76, 1537-1557.

American Cancer Society, 2014a. Cancer Facts and Figures. http://www.cancer.org/Research/CancerFactsStatistics/CancerFactsFigures2014/cancer-facts-and-figures-2014.pdf (8 February 2015, date last accessed).

American Cancer Society, 2014b. Estimated New Cancer Cases by Sex and Age. http://www.cancer.org/Research/CancerFactsStatistics/CancerFactsFigures2014/estimated-new-cases-site-sex-age-2014.pdf (8 February 2015, date last accessed).

American Society of Clinical Oncology, 2015a. Aging and Cancer. http://www.cancer.net/navigating-cancer-care/older-adults/aging-and-cancer (8 February 2015, date last accessed).

American Society of Clinical Oncology, 2015b. The Genetics of Cancer. http://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genetics-cancer (8 February 2015, date last accessed).

Anderson, K.H., Burkhauser, R.V., 1985. The Retirement-Health Nexus: A New Measure of an Old Puzzle. Journal of Human Resources 20, 315-330.

Angrist, J., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, New Jersey.

Armitage, P., Doll, R., 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 8, 1-12.

Bailey, M., Hershbein, B., Miller, A., 2012. The Opt-In Revolution? Contraception and the Gender Gap in Wages. American Economic Journal: Applied Economics 4, 225-254.

Becker, S.O., Ichino, A., 2002. Estimation of average treatment effects based on propensity scores. The Stata Journal 2, 358–377.

Black, D., Smith, J., 2004. How robust is the evidence on the effects of college quality? Evidence from matching. Journal of Econometrics 121, 99-124.

Bound, J., 1991. Self-Reported versus Objective Measures of Health in Retirement Models. Journal of Human Resources 26, 106-138.

Bradley, C.J., Neumark, D., Zhehui, L., Bednarek, H., Schenk, M., 2005. Employment Outcomes of Men Treated for Prostate Cancer. J Natl Cancer I 97, 958-965.

Bradley, C.J., Neumark, D., Zhehui, L., Schenk, M., 2007. Employment and cancer: findings from a longitudinal study of breast and prostate cancer survivors. Cancer Invest 25, 47-54.

Bureau of Labor Statistics, 2013. Labor force projections to 2022: the labor force participation rate continues to fall. http://www.bls.gov/opub/mlr/2013/article/labor-force-projections-to-2022-the-labor-force-participation-rate-continues-to-fall.htm (8 February 2015, date last accessed).

Danaei, G., Vander Hoorn, S., Lopez, A.D., Murray, C.J., Ezzati, M., et al., 2005. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. Lancet 366, 1784-1793.

Duggan, M., Singleton, P., Song, J., 2007. Aching to retire? The rise in the full retirement age and its impact on the social security disability rolls. Journal of Public Economics 91, 1327-1350.

Finkelstein, A., 2004. The interaction of partial public insurance programs and residual private insurance markets: evidence from the US Medicare program. Journal of Health Economics 23, 1-24.

Frölich, M., 2004. Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators. Review of Economics and Statistics 86, 77-90.

Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing Selection Bias Using Experimental Data. Econometrica 66, 1017-1098.

Heckman, J., Ichimura, H., Todd, P., 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies 64, 605-654.

Ichino, A., Mealli, F., Nannicini, T., 2008. From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? Journal of Applied Econometrics 23, 305 - 327.

Moran, J.R., Short, P.F., Hollenbeak, C.S., 2011. Long-term employment effects of surviving cancer. Journal of Health Economics 30, 505-514.

Nannicini, T., 2007. Simulation-based sensitivity analysis for matching estimators. The Stata Journal 7, 334 - 350.

Nordling, C.O., 1953. A new theory on cancer-inducing mechanism. Br J Cancer 7, 68-72.

Odden, M.C., Shlipak, M.G., Whitson, H.E., Katz, R., Kearney, P.M., defilippi, C., et al., 2014. Risk factors for cardiovascular disease across the spectrum of older age: the Cardiovascular Health Study. Atherosclerosis 237, 336-342.

Rothman, K.J., Poole, C., 1988. A strengthening programme for weak associations. Int J Epidemiol 17, 955-959.

Rottenberg, Y., Jacobs, J., Barchana, M., Stessman, J., 2011. Risk factors for cancer among elderly: The Jerusalem longitudinal cohort study. Journal of Geriatric Oncology 2, 45-49.

Sheather, S.J., 2004. Density Estimation. Statistical Science 19, 588-597.

Short, P.F., Vasey, J.J., Moran, J.R., 2008. Long-Term Effects of Cancer Survivorship on the Employment of Older Workers. Health Services Research 43, 193-210.

Surveillance, Epidemiology, and End Results Program, 2015. Fast Statistics. http://seer.cancer.gov/faststats/selections.php?series=age (8 February 2015, date last accessed).

Sweeney, C., Blair, C.K., Anderson, K.E., Lazovich, D., Folsom, A.R., 2004. Risk factors for breast cancer in elderly women. Am J Epidemiol 160, 868-875.

Wittenburg, D.C., Stapleton, D.C., Scrivner, S.B., 2000. How Raising the Age of Eligibility for Social Security and Medicare Might Affect the Disability Insurance and Medicare Programs. Social Security Bulletin 63, 17-26.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP14/05 Sandra E Black, Paul J Devereux and Kjell G Salvanes: 'Does Grief Transfer across Generations? - In-Utero Deaths and Child Outcomes' March 2014
WP14/06 Morgan Kelly and Cormac Ó Gráda: 'Debating the Little Ice Age' March 2014
WP14/07 Alan Fernihough, Cormac Ó Gráda and Brendan M Walsh: 'Mixed Marriages in Ireland A Century Ago' March 2014
WP14/08 Doireann Fitzgerald and Stefanie Haller: 'Exporters and Shocks: Dissecting the International Elasticity Puzzle' April 2014
WP14/09 David Candon: 'The Effects of Cancer in the English Labour Market' May 2014
WP14/10 Cormac Ó Gráda and Morgan Kelly: 'Speed under Sail, 1750–1850' May 2014
WP14/11 Johannes Becker and Ronald B Davies: 'A Negotiation-Based Model of Tax-Induced Transfer Pricing' July 2014
WP14/12 Vincent Hogan, Patrick Massey and Shane Massey: 'Analysing Match Attendance in the European Rugby Cup' September 2014
WP14/13 Vincent Hogan, Patrick Massey and Shane Massey: 'Competitive Balance: Results of Two Natural Experiments from Rugby Union' September 2014
WP14/14 Cormac Ó Gráda: 'Did Science Cause the Industrial Revolution?' October 2014
WP14/15 Michael Daly, Liam Delaney, Orla Doyle, Nick Fitzpatrick and Christine O'Farrelly: 'Can Early Intervention Policies Improve Well-being? Evidence from a randomized controlled trial' October 2014
WP14/16 Thérèse McDonnell and Orla Doyle: 'Maternal Employment, Childcare and Childhood Overweight during Infancy' October 2014
WP14/17 Sarah Parlane and Ying-Yi Tsai: 'Optimal Sourcing Orders under Supply Disruptions and the Strategic Use of Buffer Suppliers' October 2014
WP14/18 Morgan Kelly and Cormac Ó Gráda: 'Ready for Revolution? The English Economy before 1800' November 2014
WP14/19 Johannes Becker, Ronald B Davies and Gitte Jakobs: 'The Economics of Advance Pricing Agreements' November 2014
WP14/20 David Madden: 'Bridging the Gaps: Inequalities in Childrens' Educational Outcomes in Ireland ' November 2014
WP14/21 Ronald B Davies, Julien Martin, Mathieu Parenti and Farid Toubal: 'Knocking on Tax Haven's Door: Multinational Firms and Transfer Pricing' December 2014
WP15/01 Ronald B Davies, Benjamin H Liebman and Kasaundra Tomlin: 'I've Been Everywhere (Except Mexico): Investor Responses to NAFTA's Cross-Border Trucking Provisions' January 2015
WP15/02 David Madden and Michael Savage: 'Which Households Matter Most? Capturing Equity Considerations in Tax Reform via Generalised Social Marginal Welfare Weights' February 2015
WP15/03 Ronald B Davies, Rodolphe Desbordes and Anna Ray: 'Greenfield versus Merger & Acquisition FDI: Same Wine, Different Bottles?' February 2015
WP15/04 David Candon: 'Are Cancer Survivors who are Eligible for Social Security More Likely to Retire than Healthy Workers? Evidence from Difference-in-Differences' February 2015
WP15/05 Morgan Kelly and Cormac Ó Gráda: 'Adam Smith, Watch Prices, and the Industrial Revolution' March 2015
WP15/06 Kevin Denny: Has Subjective General Health Declined with the Economic Crisis? A Comparison across European Countries' March 2015