1-1-2017

Essays On Health And Labor Market Outcomes

Maryam Jafari Bidgoli
Wayne State University,

ESSAYS ON HEALTH AND LABOR MARKET OUTCOMES

by

MARYAM JAFARI BIDGOLI

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2017

MAJOR: ECONOMICS

Approved By:

Advisor Date

DEDICATION

To my loving mother and father; Fatemeh and Mahmoud,

for their endless love, support and encouragement, and

to my sisters; Mahnaz and Marjan,

who have never left my side.

ACKNOWLEDGMENTS

It is hard to believe my Ph.D. journey has come to an end, and today is the day to write this note of thanks as the last finishing touch on my dissertation. Studying a Ph.D. has been a period of intense learning for me, not only in economics, but also on a personal level. There is no doubt I could not go through this difficult phase of my life without the support that I have received from my family, friends and professors.

Foremost, I would like to express my sincere gratitude to my advisor Dr. Allen Goodman for his immeasurable amount of support and patience throughout my journey. He has been supportive since the first day I began my Ph.D. at WSU, and has supported me academically and emotionally through the rough road to finish this dissertation. My sincere appreciation is extended to my other committee members, Dr. Gail Jensen Summers and Dr. Jennifer Ward-Batts for their insightful comments and guidance that intensively helped me to improve my research. And many thanks to Dr. Janet Hankin for her time and valuable suggestions.

I am thankful to Dr. Stephen Spurr and Dr. Li Way Lee for their support, and my dear friend Elham Mahmoudi who continuously encouraged me to pursue my dreams, and complete my doctoral program.

I must express my gratitude to my patient parents, and my loving sisters who experienced all my ups and downs during the most difficult times of my Ph.D., and never stopped their moral and emotional support, and gave me a sympathetic ear.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

The amount of time a person can spend producing earnings depends on his/her stock of health as a durable human capital stock. Individuals' initial health stock depreciates with age (Grossman 1972). Health is generally considered as an important determinant of individuals' labor supply. Poor health may affect time allocation between leisure and work, and reduce the total amount of time available to spend on the labor market. Impaired health is a major cause of non-employment in middle age and older, and it is a significant constraint on the earning capacity and employment opportunities of aging populations (García-Gómez et al. 2013). According to the World Health Organization (WHO), chronic diseases are a major cause of mortality globally, and have significant effects on people's physical activities. We cannot neglect the economic burden of chronic diseases on individuals' lives in terms of non-employment and early retirement. Simultaneously, technological advances in medical treatments have caused people to live longer. The advances also affect function and quality of life of those with diseases. Therefore, individuals suffering from chronic diseases are now more likely to remain in the labor market. Evaluating the economic and social burdens of such health impairments is essential.

The main focus of this dissertation is the association between health on labor market outcomes. However, we test other relevant hypotheses including the effect of health insurance, incident of cancer on the probability of working as well as the hours of work, and the duration of non-employment. We consider the Health and Retirement Study (HRS) and RAND HRS for a large, representative national US sample of respondents who are below 65. The panel structure of the data allows for following up individuals every two years from the age 51 with individual health, labor market status, financial and socio-economic explanatory variables, and many other details. We use a pooled sample of eight waves (1996-2010), and conduct analysis for males and females. A large, longitudinal data set allows for reducing the effect of endogeneity bias, and the estimates are more efficient compared with cross-sectional analyses (Cai 2010; Christensen and Kallestrup-Lamb 2012).

The rest of the dissertation is laid out as follows. Chapter 2 studies the relationship between health and labour supply, and treats health as an endogenous variable by employing a simultanous equation approch. Two broad method of estimations has been used; Partial and Full Information Maximum Likelihood (PIML and FIML). Under PIML, we conduct two-stage prediction substitution (2SPS) and two-stage residual inclusion (2SRI). We also conduct true endogeneity test of healh by using the estimates from the FIML method. Chapter 3 examines the effect of health insurance on labor supply assuming employer provided health insurance is endogenous with respect to labor supply. Chapter 4 studies the short- and long-term impacts of cancer on labor supply and hours of work. It also examines the effect of spouse's health insurance and spouse's earning on cancer survivors' labor supply compared with a non-cancer group. Chapter 5 examines the effect of incidence of cancer on the duration of non-employment. Chapter 6 summarizes all of the results, and considers possible extensions.

CHAPTER 2. HEALTH AND LABOR SUPPLY: A SIMULTANOUS EQUATION MODEL

2.1. Literature reviews

Many studies have focused on the linkage between poor health and labor market outcomes. The impact of impaired health on labor supply has previously been analyzed using various proxies for poor health. Others research has focused instead on the effect of labor outcomes, such as wage and hours of work, on health. Few studies have considered health as an endogenous variable and simultaneously determined the effect of health on labor supply and vice versa.

Many studies have treated health as an exogenous variable, and used different ways to measure and include health in the labor supply equation. Some of them included health in the labor supply equation using a discrete self-reported health status (poor, fair, good, very good, and excellent) variable. Others have narrowed their focus to a specific disease such as arthritis or cancer, or have focused on disability (Bradley et al. 2002; Jean and Burkhauser 1990; Bradley et al. 2005; Stern 1989). Using the 1978 Survey of disability and Work (SDW) age 18 to 64, Jean and Burkhauser (1990) studied the effect of poor health on both wage rates and hours of work. They used a simultaneous Tobit model for hourly wage and hours worked to examine the impact of arthritis on wages and hours of work. They argue that arthritis ideal for studying the effect of poor health on labor market activities in the sense that it is the most common chronic disease and also the second leading cause of work disability in the US. They found that the total wage earnings of those suffering from arthritis are significantly below those of healthy workers. Bradley et al. (2002) examined the effect of breast cancer on women's labor supply. They estimated the probability of working for a group of women who have had breast cancer. Using the 1992 health and Retirement Study (HRS), they found that the probability of working is 10 percentage points lower for breast cancer survivors than women without cancer.

If (reported) health is endogenous with respect to labor supply, then including health as an exogenous factor in modeling labor supply will cause the estimated effect to be biased. Few researches (Stern 1989; Cai and Kalb 2006; Cai 2010) have tried to address the potential endogeneity of health in the labor supply equation. When the measure of health is based on respondents' self-reports, researchers are more concerned about this bias than when more objective health measures are used. For example, some people may underrate their health to justify their non-employment status. Thus, including health as an exogenous variable may result in an upward biased effect of health on labor supply.

The endogeneity of health has been addressed by measuring health differently or by using different econometric approaches. Using the National Longitudinal Survey of Men (NLSM), Lee (1982) examined the relationship between health and wage using a structural equations model. Lee found that the wage rate coefficient in the health equation is significantly positive, and also that the health coefficient in the wage equation is significantly positive. After correcting for measurement error in self-reported health, the effect of health on wage is still strong and positive, but about 28 percent lower then the uncorrected estimate. He concluded that wages and health capital are significantly jointly determined.

Using a simultaneous equations model of labor force participation and endogenous self-reported disability, Stern (1989) found that participation is statistically insignificant in the disability equation, and that disability measures are all statistically significant in the labor force participation equation. He addresses two potential sources of disability endogeneity: a direct effect of participation on disability, such as the effect of poor working conditions, and errors in self-reports of disability. Cai and Kalb (2006) follow Stern's approach, and using the Household, Income and Labor Dynamics in Australia (HILDA) survey, they conduct a simultaneous equation of health and labor force participation. They estimated the effect of self-assessed health on participation, and found health to be endogenous in the labor force participation equation. In a separate study, Cai (2010) used Australian panel data to estimate the effect of health on labor force participation using the same method. His findings showed that health has a positive and significant effect on labor force participation, and that labor force participation has a negative effect on health for males, but a positive effect on health for females.

Among the studies focused on specific chronic conditions, most treated the incidence of chronic diseases as exogenous (Bradley et al. 2013; Bradley et al. 2012; Bradley et al. 2007; Bradley et al. 2005). Zhang et al. (2009) examined the effect of health on labor force participation by including the incidence of chronic diseases. Their finding rejected exogeneity of chronic diseases. Nevertheless, it has been argued that use of specific chronic conditions reduces the potential measurement error as compared to using self-reported health status.

In this dissertation we estimate the effect of health on labor supply for males and females, treating health as an endogenous variable. Following Stern (1989) and Cai (2010), we use a simultaneous equations model to take into account potential endogeneity of health with respect to labor supply. We measure health using a subjective self-report. We also include measures of physical function and chronic illnesses in the health equation. Physical functionality is measured by Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs). We use principal component analysis to create a single continuous variable from multiple measures.

Unlike Stern (1989) and Cai (2010), we allow for health insurance coverage to affect labor supply. Since Australians have access to public health care, the crucial role of health insurance in labor market outcomes was not accounted for in Cai's (2010) study. It is likely that health insurance coverage is endogenous with respect to labor supply. To address this, we rerun our model with a limited sample of married males and females who have health insurance through their spouses' employer. Unlike Stern (1989), we use pooled panel data that allows us to control for heterogeneity, so our estimates are more efficient than a cross-sectional data analysis (Cai 2010).

We employ two estimation methods: partial (PIML) and full (FIML) information maximum likelihood. Terza et al. (2008) assert that the two-stage residual inclusion (2SRI) estimator is generally consistent, while the two-stage predictor substitution (2SPS) method is not. Thus, unlike Stern (1989) and Cai (2010), we employ both 2SPS and 2SRI methods in our PIML model. We use a Conditional Mixed Process estimator (CMP) with multilevel random effects and coefficients to conduct a FIML estimation (Roodman 2011).

The first chapter of this dissertation provide an answer to the following question: Testing the so-called justification hypotheses: what is the impact of self-reported health on the individuals' labor supplies taking into account the endogeneity of health with respect to labor supply? i.e. whether individuals justify their non-employment status by reporting poor health.

2.2. Theoretical Framework

This section presents a simultaneous equations model of labor supply and health accounting for endogeneity of health. We follow the theoretical framework of Stern (1989) and Cai (2010).

The variation in the value of Labor Supply (LS) can be estimated by the variation in true (but unmeasured) health, and a set of exogenous variables. The first equation specifies the determination of labor supply

$$(\text{Labor Supply})_{t} = \gamma_{L}(\text{True health}) + (\text{Exogenous vars})_{L,t}\varphi_{L} + (e)_{L,t} \qquad (1)$$

γ_L and φ_L are coefficients to be estimated.

The second equation specifies latent (true) health as a function of labor supply and a set of exogenous variables.

$$(\text{True health})_t = \gamma_H(\text{Labor supply})_t + (\text{Exogenous vars})_{H,t}\varphi_H + (e)_{H,t} \qquad (2)$$

The true value of health is unobservable. Thus, we need another equation that presents the relationship between true health and observable health measures, such as self-reported health scores. The third equation represents the observed (self-reported) health status as a function of the true value of health and labor supply. The dependency of self-reported health status indicates the endogeneity of self-reported health. A positive γ_H would imply that those working for pay tend to overstate their health and those not working for pay tend to understate their health.

$$(\text{Observed health})_t = (\text{True health})_t + \gamma_H(\text{Labor supply})_t + (e)_{0,t} \qquad (3)$$

Three error terms are assumed to be jointly normally distributed.

By substituting Equation (2) into Equations (1) and (3), we obtain Equations (4) and (5).

$$(Observed\ health)_t = \theta_H(Labor\ supply)_t + (exogenous\ vars)_{H,t}\varphi_H + (e)_{H,t} \qquad (4)$$

$$where\ \theta_H = \gamma_H + \gamma_L$$

$$(labor\ supply)_t = \frac{\gamma_L}{(1+\gamma_L\gamma_H)}(Observed\ health)_t + (Exogenous\ vars)_{Lt}\frac{\varphi_L}{(1+\gamma_L\gamma_H)} + (e)_{L,t}$$

$$\text{where}\ \frac{\gamma_L}{(1+\gamma_L\gamma_H)} = \theta_L\ \text{and}\ \frac{\varphi_L}{(1+\gamma_L\gamma_H)} = \varphi_L \qquad (5)$$

θ_H is a coefficient to be estimated.

In many surveys, including the HRS, respondents are asked to rate their health from poor to excellent; poor (=1), fair (=2), good (=3), very good (=4), and excellent (=5). Thus, the observed endogenous health variable is:

$$\text{H}=\kappa\ \ \text{if } m_i < \text{unobserved health} \le m_j\ \text{ where } k = 1,2,3,4, \text{and } 5 \qquad (6)$$

$$where\ (i = -1, 0, 1, 2, and\ 3\ ,\ \ m_{-1} = -\infty)\ \ and\ (\ j = 0, 1, 2,3\ and\ 4\ \ , m_4 = +\infty)$$

We observe self-reported health status, but not cut-off points in an underlying continuous observed health measure, which are coefficients to be estimated.

The endogenous labor supply variable is:

$$\mathrm{L_t} = \begin{cases} 1\ (= working\ for\ pay) \\ 0\ (= not\ working\ for\ pay) \end{cases} \tag{7}$$

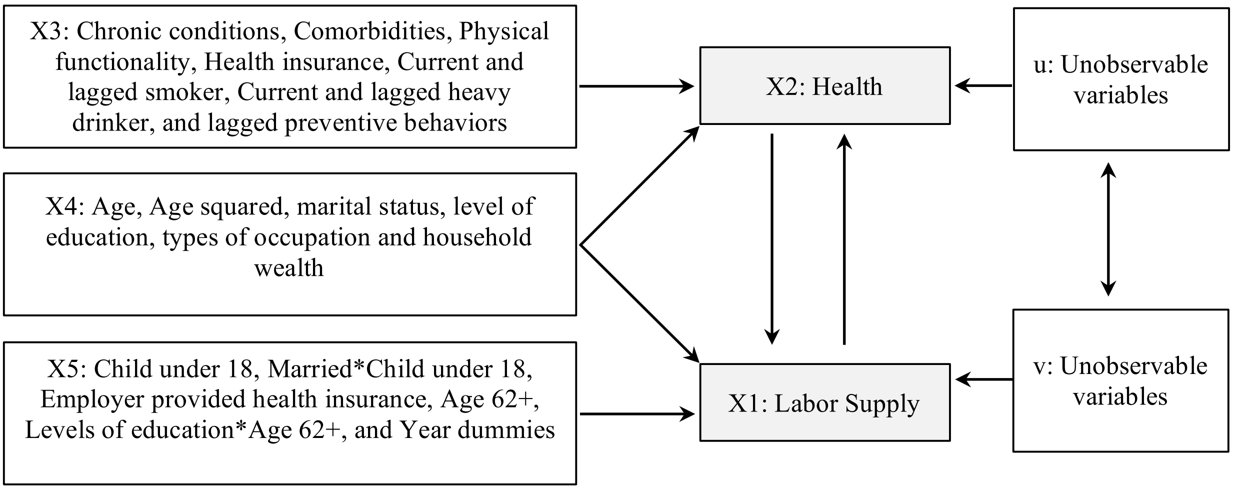
Equations 4 to 7 are used to construct a simultaneous equations model. θ_H, φ_H, θ_L, $and\ \varphi_L$ are coefficients to be estimated. In addition, $m_0, m_1, m_2, and\ m_3$ are health cut-off points to be estimated.

The modeling approach is similar to Stern (1989), who estimated the model using cross sectional US data, and Cai (2010), who uses Australian longitudinal data. We follow their method to estimate the effect of endogenous health on labor supply. We use both the two-stage predictor substitution (2SPS), and two-stage residual inclusion (2SRI) methods. As Terza et al. (2008) argued, the 2SRI estimator is generally consistent while the 2SPS method is not. The 2SRI approach was first discussed by Hausman (1978), and developed further by Smith and Blundell (1986). The 2SRI method, instead of including the predicted value of an endogenous variable from the first stage in the second stage, includes the residuals of the first stage in the second stage, while also including the observable endogenous variable as a regressor in the second stage. Terza et al. (2008) also argue that, like two-stage least squares (2SLS) for linear models, the 2SPS approach for nonlinear models is not consistent. The 2SRI method addresses this limitation.

2.3. Variables

We use the following graph as a better illustration of the relationship between health and labor supply variables as endogenous variables, and other exogenous and control variables.

Figure 1. Reciprocal causation



In this model, X1 and X2 are considered endogenous variables, which are determined in the model simultaneously. X3, X4, and X5 are sets of exogenous variables that are determined outside the model. u and v are the residuals and are correlated. The arrows from X4 to X5 and from X5 to X4 indicate the endogeneity of health and employment status.

Exclusion restrictions are required to identify the simultaneous equation model. The following paragraphs illustrate the included and excluded variables in each equation of the model, and provide definitions of the variables.

X1, Labor supply is defined as a binary variable that equals one if the respondent reports currently working for pay.

X2, Health status is the respondent's self-reported general health status, scaled from "1" for poor to "5" for excellent.

X3, is a set of exogenous variables included in the health equation and excluded from the labor supply equation: Chronic conditions, Number of chronic conditions (comorbidity) physical functionality, health insurance, lagged preventive behaviors, current smoker, lagged smoker, current heavy drinker[1], lagged heavy drinker are the included variables in the health equation. The chronic conditions are high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, psychiatric problems, and arthritis. Respondents were asked whether or not a doctor told the respondent he/she had each condition.

We use ADLs (five tasks of bathing, eating, dressing, walking across a room, and getting in or out of bed) and IADLs (using a telephone, talking meditation, handling money, shopping, preparing meals) to construct a physical functionality variable. Following (Ginneken and Groenewold 2012), we use Principal Component Analysis (PCA) of 10 items to create a single index of physical functionality. The first component explains the most of the variance. According to the eigenvalues, we should retain two components because they have eigenvalues over 1.00 explaining over 50% of the variance cumulatively. However, in terms of the explained variation by the components, the second component does not contribute as much as the first component. In other words, there is a significant break between two components. This leads us to the conclusion that a one factor solution will probably be adequate. The conclusion is supported by scree plot of eigenvalues after PCA. PCA results are reported in the Appendix A. As Cai (2010) argues, chronic health conditions and physical functioning may be treated as exogenous variables. Although these are also reported by the respondents, they are less subjective. It is difficult to estimate the effect of health insurance on health because the same determinants are expected to influence both health and health insurance coverage. In addition, health status may directly affect insurance coverage (Levy and Meltzer 2008).

X4, is a set of variables that affects both health and labor supply. These include age, age squared, marital status (married versus unmarried), level of education (less than high school completion, high school, some college/ college degree, more than college), type of

[1] According to National Institute on Alcohol Abuse and Alcoholism, Heavy drinking defines as drinking 5 or more drinks on the same occasion on each of 5 or more days in the past 30 days. We define a heavy drinker as a man who drinks more than 5 (4) standards drinks per day when drinking or drinks five days a week.

occupation (white collar 1, white collar 2, and blue collar)[2], household wealth[3]. The potential wage is also a factor that influences both health and labor supply equations. We include level of education, type of occupation, age, and age squared as proxies for the potential wage in both equations.

X5 is a set of variables that are included in the labor equation and excluded from the health equation. Having a young child (under 18) residing with the respondent is an obstacle to labor supply. The interaction between the presence of a resident child under 18 and marital status is also included in the labor supply equation. As previous studies argue, employer-provided health insurance is endogenous to the labor supply (Bradley et al. 2013; Bradley et al. 2012). We first estimate the model assuming employer provided health insurance is exogenous. Then in chapter 3, we will rerun the model for a group of married individuals who have health insurance through their spouse to take the endogeneity of the health insurance into account. We also include age over 62 as a dummy regressor in the labor supply equation because an individual can choose to retire as early as age 62. Also, the interaction between age over 62 and level of education, year dummies, current smoker, current heavy drinker, and lagged preventive behaviors[4] are included in the labor supply equation.

2.4. Econometric Approach

2.4.1. Limited Information Maximum Likelihood Method (LIML)

We employ a two-step nonlinear estimator to estimate health and employment status, allowing for endogeneity of these variables. The reduced forms for Equation (6) and Equation

---

[2] White collar 1 includes managerial specialty operation or technical support, white collar 2 includes sales, clerical, administrative support or services, and blue collar includes farming, forestry, fishing, mechanics and repair, construction trade and extractors, precision production or operators.

[3] The net value of total wealth (excluding second home) is calculated as the sum of all wealth components less all debt.

[4] The preventive behavior is defined as whether the respondent reports preventive health tests and procedures such as a blood test for cholesterol, a flu shot, monthly self-checks for breast lumps, a mammogram, a pap smear, and a check for prostate cancer.

(7) are as follows:

$$(\mathrm{E})_{\mathrm{t}} = \mathrm{X}\Pi_{\mathrm{L}} + (\mathrm{e})_{\mathrm{L}} \tag{8}$$

$$(\mathrm{H})_{\mathrm{t}} = \mathrm{X}\Pi_{\mathrm{H}} + (\mathrm{e})_{\mathrm{H}} \tag{9}$$

We estimate Π_{L} and Π_{H} using two instrumental variables based approaches: two-stage predictor substitution (2SPS), and two-stage residual inclusion (2SRI). We compute the predicted value of employment status using a random effects probit model, and the predicted value of health using an ordered probit model.

Under the 2SPS method, we regress endogenous variables on all exogenous variables and covariates in the first stage, and obtain the predicted value of endogenous variable. In the second stage, this predicted value replaces the observed value (Stern 1989; Terza et al. 2008; Cai et al. 2011). The disadvantage of this method is that the correlation between the two equations is not taken into account (Cai 2010). Π_{L} is estimated using a probit model for panel data , and Π_{H} is estimated using ordered probit. Then we have Equation 10 and Equation 11.

$$\left(\hat{E}\right)_{\mathrm{t}} = \mathrm{X}\widehat{\Pi}_{\mathrm{L}} \tag{10}$$

$$\left(\widehat{H}\right)_{\mathrm{t}} = \mathrm{X}\widehat{\Pi}_{H} \tag{11}$$

In the second-stage, we substitute $\widehat{\Pi}_{\mathrm{L}}$and $\widehat{\Pi}_{\mathrm{H}}$ in Equation (8) and Equation (9).

Terza et al. (2008) demonstrated the superiority of the 2SRI method to the 2SPS method when try to address endogeneity in non-linear models. In the 2SRI approach, the first-stage is identical to the 2SPS. However, instead of using the predicted value of endogenous variable from the first-stage regression in the second-stage regression, we use the first-stage residuals in second-stage estimation. The observable endogenous variable is also a regressor in the second-stage equation.

We define the residuals in the model as generalized residuals shown in Equation (12). We follow Vella (1993) to calculate the generalized residuals in ordered probit model, which take the following form,

$$E(\upsilon_{ji}|d_{ji}) = \frac{d_{ji}\hat{\pi}}{\hat{\Pi}_{ji}(1-\hat{\Pi}_{ji})}(d_{ji} - \hat{\Pi}_{ji}) \tag{12}$$

where $d_{ji}$is an indicator function taking the value 1 if individual i is in category j and 0 otherwise, $\hat{\Pi}_{ji}$ is the estimated probability that individual i is in the j^{th} category, and $\hat{\pi}$ is the estimated value of the density at that point. For a probit model, the generalized residuals are calculated as follows:

$$\begin{cases} y = 1 \quad : pdf(xb)/cdf(xb) \\ y = 0: \ -pdf(xb)/(1 - cdf(xb)) \end{cases} \tag{13}$$

which is same as the inverse mills ratio for the probit model.

We include the generalized residuals, collected from the first stage health equation and add to the second-stage estimation of the labor supply equation as a regressor while retaining the observed H_t in the regression (not the predicted value of health ($\hat{H}_t$)), and also add the generalized residuals from the first stage labor supply equation and add as a regressor to the second-stage estimation of the health equation while retaining the observed E_t in the regression (not the predicted value of labor supply ($\hat{E}_t$)).

2.4.2. Full Information Likelihood Maximization method (FIML)

Using the FIML method allows for the correlation between the two equations that is not taken into account in the two-stage methods. Therefore, the two-stage methods are inefficient. In addition, applying the FIML method allows for doing the true exogeneity test of the endogenous variable (Cai 2010). We use a Conditional Mixed Process (CMP) estimator that employs a full-information maximum likelihood (FIML) method to estimate the system of equations. It has been shown by Roodman (2011) that -CMP- estimator can be used for two types of estimation. 1) A recursive data-generating process that in this case -CMP- is a FIML estimator. 2) In case of simultaneity such as 2SLS model that -CMP- is LIML estimator.

To conduct the true endogeneity test for health and also the justification test, we must model the system of equations in a way that we can estimate the variance-covariance of error terms of two equations. Borrowing the approach has been used by Cai (2010) for four waves of data which are used his paper, we extend his approach readily to our data with eight waves. As he has discussed, given the covariance between $(\varepsilon)_{h,t}$and $(\varepsilon)_{L,t}$, we can construct the variance-covariance matrix.

$$cov(\varepsilon_{i,s},\varepsilon_{j,t}) = c\begin{cases} \delta_{i(\mu)} + \delta_{i(\nu)} & if\ i = j\ and\ s = t \\ \delta_{i(\mu)} & if\ i = j\ and\ s \neq t \\ \delta_{hl(\mu)} + \delta_{hl(\nu)} & if\ i \neq j\ and\ s = t \\ \delta_{hl(\mu)} & if\ i \neq j\ and\ s \neq t \end{cases} \quad (14)$$

for $i, j = h\ , L;\ and\ s, t = 1, \dots T$

Where $\delta_{i(\mu)} and\ \delta_{i(\nu)}$are the variances of the time-invariant and time-variant error components respectively; $\delta_{hl(\mu)}$is the covariance of the two time-invariant error components; and $\delta_{hl(\nu)}$is the covariance of the two time-variant error components.

The variance-covariance matrix for the structural model containing two Equations (4) and (5) is, $\delta_{hl(\mu)}$

$$cov(\varepsilon_{h,1}\varepsilon_{h,2},\varepsilon_{h,3},\varepsilon_{h,4},\varepsilon_{h,5},\varepsilon_{h,6},\varepsilon_{h,7},\varepsilon_{h,8};\varepsilon_{L,1},\varepsilon_{L,2},\varepsilon_{L,3},\varepsilon_{L,4},\varepsilon_{L,5},\varepsilon_{L,6},\varepsilon_{L,7},\varepsilon_{L,8})$$

$$\equiv \Omega = \begin{cases} \mathrm{I_8\delta_{h(\mu)} + e_8e_4'\delta_{h(\nu)}} & \mathrm{I_8\delta_{hL(\mu)} + e_8e_8'\delta_{hL(\nu)}} \\ \mathrm{I_8\delta_{hL(\mu)} + e_8e_8'\delta_{hL(\nu)}} & \mathrm{I_8\delta_{L(\mu)} + e_8e_8'\delta_{L(\nu)}} \end{cases} \quad (15)$$

Where $\mathrm{I_8}$ is a eight-dimensional identity matrix, and $\mathrm{e_8}$ is a column vector with eight ones as its elements. Then the covariance matrix of the reduced form is,

$$cov(\varepsilon^*_{h,1},\varepsilon^*_{h,2},\varepsilon^*_{h,3},\varepsilon^*_{h,4},\varepsilon^*_{h,5},\varepsilon^*_{h,6},\varepsilon^*_{h,7},\varepsilon^*_{L,8};\varepsilon^*_{L,1},\varepsilon^*_{L,2},\varepsilon^*_{L,3},\varepsilon^*_{L,4},\varepsilon^*_{L,5},\varepsilon^*_{L,6},\varepsilon^*_{L,7},\varepsilon^*_{L,8}) =$$

$$\Omega^* = A\Omega A' \text{ where } A = \frac{1}{1-\theta_1\theta_2}\begin{pmatrix} I_8 & I_8\theta_1 \\ I_8\theta_2 & I_8 \end{pmatrix} \quad (16)$$

The system can be estimated using Maximum Likelihood Estimation. In order to implement the FIML method, $\mu_h, \mu_{L,}\upsilon_{h,t}$ and $\upsilon_{L,t}$ are normally distributed with mean zero and variance-covariance equal to Equation 16.

Following Cia (2010), we employ Maximum Simulated Likelihood (MSL) technique. More detailed are provided in the Appendix B. MSL estimation functions based on an unbiased simulator for the likelihood function. The -CMP- estimator in Stata[5] enables us to follow Cai (2010) and use Geweke-Hajivassiliu-Keane (GHK) simulator. It has been argued by Hajivassilio and Ruud (1994) MSL estimator is consistent if the number of replication as N→ ∞. We also use antithetic acceleration in simulating the random draws to reduce variance.

2.5. Data

The data come from the RAND HRS Data file from 1996 through 2010 (eight waves), which is a cleaned and easy-to-use version of data from eleven waves of the HRS data. We excluded respondents who were non-responsive even for a single wave. The 1992 and 1994 waves were excluded from the analysis due to inconsistent question wording over time. After limiting the sample to those aged less than 65, the total number of observations (person-waves) is 62,779, consisting of 25,027 male and 37,752 female observations. The total number of individuals is 20,519: 8,749 males and 11,770 females.

Table 1 shows the total number of observations in the pooled sample of eight waves by gender and year.

Table 1. Total number of observations by gender and year

Year	Male		Female	
	(N=25,027)	%	(N=37,752)	%
1996	3,582	14.31	5,255	13.92
1998	4,003	15.99	5,702	15.10
2000	3,263	13.04	4,844	12.83
2002	2,568	10.26	4,063	10.76
2004	3,234	12.92	4,900	12.98
2006	2,426	9.69	3,968	10.51
2008	2,022	8.08	3,291	8.72
2010	3,929	15.70	5,729	15.18

Table 2 shows the total number of respondents in the pooled sample of eight waves by gender and year.

[5] Data Analysis and Statistical Software

Table 2. Total number of respondents by gender and year

Year	Male		Female	
	(N=8,749)	%	(N=11,770)	%
1996	3,582	40.94	5,255	44.65
1998	1,272	14.54	1,445	12.28
2000	92	1.05	150	1.27
2002	61	0.70	114	0.97
2004	1,359	15.53	1,656	14.07
2006	54	0.62	74	0.63
2008	49	0.56	48	0.41
2010	2,280	26.06	3,028	25.73

2.6. Descriptive Statistics

Table 3 presents descriptive statistics in the pooled eight-waves sample for males and females. The HRS is a nationally representative sample of those aged 51 and older, but spouses are included in the data regardless of age. Our sample is restricted to males aged between 22 and 64 years old and females aged between 23 and 64 years old. Our male sample is predominantly middle aged (mean age is 57.78), white (85 percent), married (81 percent), having health insurance (88 percent), mostly covered by their own employers (57 percent), and are employed (70 percent), have high school diploma or more (81.6%). Forty-one percent live in the South. The sample of females is also predominantly middle aged (mean age is 56.48), white (82 percent), married (70 percent), have health insurance (86 percent – 40 percent mostly covered by their own employers), are employed (58 percent), and have high school diploma or more (80.8 percent). Forty-three percent live in the South.

Table 3. Variable definitions and variable means.

		Male	Female
Health status	1=poor, 2=fair, 3=good 4=very good, 5=excellent	3.3637	3.4009
Physical functionality	Created using ADL & IADL	-0.4791	-0.3552
Age	Age at the middle of survey	58.1130	56.7432
		(4.4080)	(5.4108)
Married	Married; 1=yes 0=No	0.8186	0.7139
Child under 18	Has child (ren) under 18; 1=yes 0=no	0.1124	0.0856
Married & Child18	Married & having child under 18	0.1006	0.0683
> High school	Less than high school diploma	0.1592	0.1415
College	College degree/some college	0.2449	0.2665
< College	More than college	0.2738	0.2129
Age 62+	Age over 62	0.2570	0.2126
> High school & Age 62+	Less than high school & age over 62	0.0501	0.0383
College & Age 62+	College degree/some college & age over 62	0.0545	0.0500
< College & Age 62+	More than college & age over 62	0.0657	0.0377
White collar 2	White collar 2 occupation=1	0.2057	0.5459
Blue collar	Blue collar occupation=1	0.4627	0.1324
Wealth	Total household asset/100,000	4.0833	3.7071
		(11.7377)	(9.9391)
Hispanic	Hispanic=1; otherwise=0	0.0972	0.0824
African American	African American=1; otherwise=0	0.1311	0.1509
Current smoker	Current smoker=1; otherwise=0	0.6894	0.5307
Current heavy drinker	Current heavy drinker=1; otherwise=0	0.0622	0.0113
Lagged smoker	Lagged smoker=1	0.4989	0.4031
Lagged heavy drinker	Lagged heavy drinker=1	0.0471	0.0091
Lagged preventive behavior	Lagged preventive behavior=1	0.3915	0.4561
Midwest	Living in Midwest	0.2598	0.2547
Northeast	Living in Northeast	0.1470	0.1499
West	Living in West=1	0.1937	0.1837
Chronic condition	Number of chronic conditions	1.3779	1.4349
		(1.2661)	(1.2949)
High blood pressure	High blood pressure; 1=yes 0=no	0.4452	0.3933
Diabetes	Diabetes; 1=yes 0=no	0.1549	0.1170
Cancer	Cancer; 1=yes 0=no	0.0584	0.0872
Lung diseases	Lung diseases; 1=yes 0=no	0.0490	0.0615
Heart diseases	Heart diseases; 1=yes 0=no	0.1682	0.1096
Stroke	Stroke; 1=yes 0=no	0.0411	0.0293
Psychiatric problem	Psychiatric problems; 1=yes 0=no	0.0912	0.1574
Employer provided HI	Health insurance; 1=yes 0=no	0.6065	0.4472
Health insurance	Health insurance; 1=yes 0=no	0.8954	0.8902
year 1998	1 if interviewed in 1998	0.1693	0.1586
year 2000	1 if interviewed in 2000	0.1381	0.1377
year 2002	1 if interviewed in 2004	0.1161	0.1212
year 2004	1 if interviewed in 2004	0.1477	0.1464
year 2006	1 if interviewed in 2006	0.1114	0.1188
year 2008	1 if interviewed in 2008	0.0921	0.0961
year 2010	1 if interviewed in 2010	0.0772	0.0797

Means or sample percentages are reported with standard deviations of continuous variables in parenthesis. Reference groups are unmarried, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996.

Table 4 shows the total number of new respondents for the sample of pooled eight waves by year and gender.

Table 4. Number of new respondents by gender and year.

	Male			Female			
Year	No.	%	Total	No.	%	Total	Total
1996	2577	100.00%	2577	3792	100.00%	3792	6369
1998	939	31.88%	2945	1102	25.94%	4249	7194
2000	67	2.79%	2401	112	3.03%	3691	6092
2002	47	2.33%	2019	82	2.53%	3247	5266
2004	1132	44.08%	2568	1342	34.21%	3923	6491
2006	0	0.00%	1937	1	0.03%	3184	5121
2008	0	0.00%	1602	0	0.00%	2575	4177
2010	0	0.00%	1342	1	0.05%	2137	3479
Observation	4762	27.38%	17391	6432	24.00%	26798	44189

Table 5 tabulates employment status against self-reported health status using the pooled sample.

Table 5. Labor supply status by self-reported health.

Employment status	Health status					
	Poor (1)	Fair (2)	good (3)	Very good (4)	Excellent (5)	All
Male						
% non-employment	80.90	52.41	33.78	27.93	25.95	35.93
% employment	19.10	47.59	66.22	72.07	74.05	64.07
Observations	1,429	4,154	7,791	9,093	4,331	26,798
Female						
% non-employment	74.49	44.61	27.34	20.58	14.31	28.54
% employment	25.51	55.39	72.66	79.42	85.69	71.46
Observations	1,031	2,634	5,424	5,583	2,719	17,391

As Table 5 shows there is a positive relationship between employment status and health status for both males and females. In other words, the better the health, the more likely to be employed.

2.7. Endogeneity test of health

To test the endogeneity of health to the labor supply, three methods are used. We test the endogeneity of health for the sample of males and females separately.

1). Assuming ρ=0, we test the significance of coefficient θ_H from the 2SPS method in *Eq. 4*. The result of test indicates that health is endogenous to the labor supply for both males and females. This method assumes that the correlation between two equations is zero, and this is the disadvantage of this method (Stern 1989).

2). We use an augmented Hausman test that was first proposed by Hausman (1978), and then developed by Smith and Blundell (1986). We use 2SRI results and add the first-stage residuals to the second-stage of the labor supply regression as an exogenous regressor. If the coefficient of the added regressor is significant, then exogeneity is rejected. The result for the sample of males indicates that health is endogenous with respect to labor supply for both males and females. Both methods assume that the correlation between two equations are zero (ρ=0). Therefore, they are only partial tests for endogeneity of health.

3). To conduct a true test of exogeneity, we follow Cai (2010), and use the FIML estimation results to measure the joint significant of θ_H and ρ. We test the following hypothesis.

$$\begin{cases} \text{H0}: \ \theta_H = 0, \delta_{hL(\mu)} = 0, \ \text{and}\ \delta_{hL(\nu)} = 0 \\ \text{H1}: \ \theta_H \neq 0, \delta_{hL(\mu)} \neq 0, \ \text{and}\ \delta_{hL(\nu)} \neq 0 \end{cases}$$

where θ_H, $\delta_{hL(\mu)}$ and $\delta_{hL(\nu)}$ are the coefficient on the labor supply variable, the covariance of the time-invariant error component, and the correlation coefficient of time-variant error components respectively. The test statistic is significant for both males and females, implying that health should not be treated as exogenous to labor supply.

Table 6. Endogeneity test using different methods of estimation.

Methods	Hypotheses
(1)	H0: $\theta_H = 0$ H1: $\theta_H \neq 0$
(2)	H0: $1st - stage\ rediduals = 0$ H1: $1st - stage\ rediduals \neq 0$
(3)	H0 : $\theta_H = 0, \delta_{hL(\mu)} = 0$, and $\delta_{hL(\nu)} = 0$ H1 : $\theta_H \neq 0, \delta_{hL(\mu)} \neq 0$, and $\delta_{hL(\nu)} \neq 0$

The different tests are summarized in Table 6.

2.8. Results

2.8.1. 2SPS and FIML estimation methods

First, we follow Cai (2010), and focus on 2SPS two-stage and FIML estimation methods. Table 7 provides the results of two-stage model of 2SPS and FIML for males and females. In section 2.8.2, we will discuss 2SRI two-stage estimation results.

The results of first-stage estimations are reported in APPENDIX C.

Table 7. Coefficient estimates and estimates of the variance-covariance parameters, 2SPS and FIML.

	Male (N=17,391)		Female (N=26,798)	
	2SPS	FIML	2SPS	FIML
Labor supply equation				
Health	0.8682***	0.6833***	0.6890***	0.5953***
	(0.043)	(0.017)	(0.030)	(0.016)
Age	0.1234	0.2000	0.3985***	0.3328***
	(0.166)	(0.167)	(0.075)	(0.066)
Age squared	-0.0022	-0.0028*	-0.0046***	-0.0039***
	(0.002)	(0.001)	(0.001)	(0.001)
Married	0.2396***	0.1900***	-0.3424***	-0.3037***
	(0.083)	(0.069)	(0.057)	(0.051)
Child 0-18	0.1102	0.2738	0.1910	0.1154
	(0.309)	(0.190)	(0.214)	(0.179)
Married*Child 0-18	-0.0335	-0.1930	-0.3752	-0.2835
	(0.331)	(0.203)	(0.235)	(0.197)
> High school	0.3863***	0.2441***	0.0018	0.0072
	(0.120)	(0.090)	(0.090)	(0.079)
College/some college	-0.0921	-0.0597	0.0069	0.0106
	(0.104)	(0.081)	(0.073)	(0.065)
< College	-0.0716	-0.0182	-0.3293***	-0.2708***
	(0.120)	(0.092)	(0.091)	(0.079)
Age 62+	-0.5285***	-0.3060***	-0.2130*	-0.1768*
	(0.148)	(0.091)	(0.110)	(0.092)
Less than high school*Age 62+	-0.0610	0.0705	0.2045	0.1659
	(0.180)	(0.111)	(0.151)	(0.126)
College*Age 62+	0.1590	0.1179	0.0036	-0.0098
	(0.172)	(0.106)	(0.135)	(0.113)
More than college*Age 62+	0.2123	0.1441	0.0368	-0.0235
	(0.170)	(0.105)	(0.148)	(0.125)
White collar 2 occupation	0.1112	0.0875	-0.0927	-0.0768
	(0.095)	(0.079)	(0.063)	(0.057)
Blue collar occupation	-0.1492*	-0.1151	-0.0958	-0.0796
	(0.088)	(0.074)	(0.089)	(0.080)
Wealth [a]	-0.0021	-0.0017	-0.0092***	-0.0081***
	(0.003)	(0.002)	(0.002)	(0.002)
Hispanic	0.4417***	0.3595***	0.4569***	0.3942***
	(0.113)	(0.094)	(0.091)	(0.081)
African American	-0.0190	-0.0268	0.2625***	0.2276***

	(0.093)	(0.077)	(0.070)	(0.062)
Midwest	0.0691	0.0518	0.1418**	0.1181**
	(0.079)	(0.066)	(0.061)	(0.055)
Northeast	0.0369	0.0231	0.2459***	0.2086***
	(0.094)	(0.078)	(0.072)	(0.065)
West	0.0760	0.0575	-0.0882	-0.0818
	(0.089)	(0.074)	(0.069)	(0.062)
Employer provided HI	0.7370***	0.5830***	1.3795***	1.1876***
	(0.066)	(0.043)	(0.053)	(0.048)
Year 1998	0.2286**	-0.1123	0.3018***	0.0799
	(0.102)	(0.075)	(0.081)	(0.080)
Year 2000	0.1119	-0.0291	0.0971	0.0393
	(0.110)	(0.076)	(0.086)	(0.073)
Year 2002	-0.0650	-0.1290	0.0785	-0.0010
	(0.115)	(0.079)	(0.090)	(0.078)
Year 2004	0.2303**	-0.0307	0.2305***	0.0943
	(0.115)	(0.077)	(0.087)	(0.077)
Year 2006	0.2858**	0.0801	0.2149**	0.1453*
	(0.127)	(0.089)	(0.093)	(0.080)
Year 2008	0.3899***	0.0242	0.4195***	0.2396***
	(0.130)	(0.093)	(0.100)	(0.090)
Year 2010	0.2127*	0.0090	0.2691***	0.2014**
	(0.126)	(0.088)	(0.100)	(0.085)
Constant	3.1912	0.7095	-5.7466***	-5.0355***
	(4.574)	(3.308)	(1.970)	(1.661)
Health equation				
Labor supply	0.8743***	0.2981***	0.0200	0.0305**
	(0.158)	(0.019)	(0.023)	(0.013)
Age	-0.0295	-0.1579	-0.0506	-0.0529
	(0.054)	(0.098)	(0.040)	(0.036)
Age squared	0.0002	0.0018**	0.0006	0.0006*
	(0.000)	(0.001)	(0.000)	(0.000)
Married	-0.0814*	-0.0397	0.1177***	0.1218***
	(0.049)	(0.047)	(0.037)	(0.035)
> High school	0.0153	-0.3729***	-0.4602***	-0.4607***
	(0.092)	(0.054)	(0.049)	(0.048)
College/some college	0.0251	0.1930***	0.2282***	0.2288***
	(0.060)	(0.048)	(0.040)	(0.040)
< College	0.0030	0.3537***	0.5241***	0.5279***
	(0.098)	(0.055)	(0.050)	(0.050)
White collar 2 occupation	-0.1212**	-0.1709***	-0.0672*	-0.0681*
	(0.053)	(0.053)	(0.041)	(0.040)
Blue collar occupation	-0.0453	-0.1737***	-0.2496***	-0.2476***
	(0.064)	(0.051)	(0.057)	(0.057)
Wealth	-0.0006	0.0026	0.0040**	0.0042***
	(0.002)	(0.002)	(0.002)	(0.002)
Hispanic	0.1373	-0.4553***	-0.6432***	-0.6589***
	(0.120)	(0.065)	(0.058)	(0.058)
African American	0.1673**	-0.1493***	-0.4192***	-0.4360***
	(0.085)	(0.054)	(0.045)	(0.045)
Midwest	-0.0332	0.0313	0.1046***	0.1098***
	(0.048)	(0.045)	(0.039)	(0.039)
Northeast	0.0045	0.0547	0.0713	0.0775*
	(0.054)	(0.053)	(0.046)	(0.046)

West	-0.0577	-0.0046	0.0768*	0.0861*
	(0.051)	(0.050)	(0.044)	(0.044)
Physical functionality	-0.0188	-0.2000***	-0.2428***	-0.2589***
	(0.042)	(0.012)	(0.014)	(0.011)
No. chronic conditions	0.0179	-0.3248***	-0.5006***	-0.4660***
	(0.089)	(0.029)	(0.033)	(0.031)
Current smoker	-0.0937*	-0.1399***	-0.0630	-0.0768*
	(0.052)	(0.043)	(0.047)	(0.044)
Current heavy drinker	0.0037	-0.0543	-0.0829	-0.1117
	(0.079)	(0.058)	(0.155)	(0.143)
Lagged smoker	-0.0163	-0.0415	-0.0629	-0.0666
	(0.051)	(0.047)	(0.051)	(0.048)
Lagged heavy drinker	-0.1617*	-0.0694	0.0872	0.0076
	(0.093)	(0.069)	(0.177)	(0.163)
Lagged preventive	0.0236	-0.0009	-0.0355	-0.0295
	(0.041)	(0.034)	(0.034)	(0.033)
High blood pressure	0.0143	0.0350	-0.0068	0.0191
	(0.054)	(0.040)	(0.049)	(0.045)
Diabetes	-0.0668	-0.1263***	-0.1291**	-0.1374**
	(0.066)	(0.047)	(0.060)	(0.055)
Cancer	-0.2210***	-0.2060***	-0.0593	-0.0651
	(0.085)	(0.063)	(0.063)	(0.059)
Lung disease	-0.1315	-0.2659***	-0.2825***	-0.2999***
	(0.101)	(0.068)	(0.075)	(0.069)
Heart disease	-0.1788***	-0.2383***	-0.1374**	-0.1546***
	(0.066)	(0.046)	(0.062)	(0.057)
Stroke	-0.1982*	-0.0104	0.2763***	0.1988**
	(0.107)	(0.073)	(0.099)	(0.092)
Psychiatric problems	-0.1680**	-0.1614***	-0.0148	-0.0902*
	(0.076)	(0.057)	(0.059)	(0.054)
Health insurance	0.0434	0.0196	0.1405***	0.0514
	(0.063)	(0.050)	(0.050)	(0.047)
Cut-1	-6.2217***	-6.3938**	-4.7369***	-4.8828***
	(1.484)	(3.019)	(1.055)	(0.994)
Cut-2	-4.7178***	-4.8876	-3.0328***	-3.1628***
	(1.483)	(3.019)	(1.054)	(0.993)
Cut-3	-3.1323**	-3.3165	-1.4294	-1.5565
	(1.483)	(3.019)	(1.054)	(0.993)
Cut-4	-1.4929	-1.6974	0.3676	0.2367
	(1.483)	(3.017)	(1.054)	(0.993)
Ln (δ_h)	0.0512***	0.0325	0.0995***	0.0970***
	(0.018)	(0.021)	(0.015)	(0.015)
Ln (δ_l)	0.5371***	0.3379***	0.4412***	0.3385***
	(0.029)	(0.028)	(0.023)	(0.0253)
$\delta_{hl(\mu)}$[b]		-0.7687***		-0.4000***
		(0.044)		(0.037)
$\delta_{hl(\nu)}$[c]		-1.0293***		-0.6082***
		(0.029)		(0.026)
Log likelihood health Eq.	-20139	NA	-29467	NA
Log likelihood labor supply Eq.	-6955	NA	-11234	NA
Total log likelihood	NA	-27039	NA	-40636

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Time-invariant error components covariance.
[c] Time-variant error components covariance.

Reference groups are unmarried, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996. Arthritis has been omitted because of collinearity. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

From Table 7, it appears that the results from 2SPS and FIML estimation methods are similar in terms of the direction and size of the effect. The results are also similar in terms of statistical significance for most of coefficients. We also conducted various specifications, and the results are very similar to Table 7.

First, we focus on the association between health and labor supply status, which is the main focus of this study. Independent of the method of estimation, a positive and highly significant effect of health on labor supply has been found for both males and females.

Since probit estimation is a non-linear estimate, it is more meaningful to interpret the results by using marginal effects (ME). With binary independent variables, ME measure discrete change. In other words, how much predicted probabilities change as the binary independent variable changes from 0 to 1, holding all other variables at their means.

Marginal Effect $X_k = pr(Y = 1|X, X_k = 1) - pr(Y = 1|X, X_k = 0)$

We use the reduced form of FIML estimation method to show the marginal effect of an independent variable on the probability of working, holding all other covariates at their means.

Table 8 provides the marginal effects of independent variables at their means using the results from the reduced form of FIML estimation method.

Table 8. Marginal effects of labor supply equation using estimates from the FIML method.

Change in y given unit change in x	Male	Female
Age	-0.0253***	-0.0246***
	(0.003)	(0.002)
Married	0.0427**	-0.0575***
	(0.017)	(0.010)
Child 0-18	0.0311*	-0.0229
	(0.018)	(0.021)
> High school	0.0023	-0.0570***
	(0.020)	(0.016)
College/some college	0.0311*	0.0339***
	(0.017)	(0.012)
< College	0.0713***	0.0089
	(0.017)	(0.016)
Age 62+	-0.0694***	-0.0385**
	(0.025)	(0.019)
White collar 2 occupation	-0.0073	-0.0277**
	(0.015)	(0.013)
Blue collar occupation	-0.0670***	-0.0545***
	(0.016)	(0.018)
Wealth [a]	0.0000	-0.0013***
	(0.001)	(0.000)
Hispanic	0.0136	0.0005
	(0.020)	(0.018)
African American	-0.0391**	-0.0076
	(0.019)	(0.014)
Midwest	0.0210	0.0434***
	(0.014)	(0.012)
Northeast	0.0175	0.0596***
	(0.017)	(0.014)
West	0.0158	-0.0075
	(0.016)	(0.014)
Employer provided health insurance	0.1802***	0.2977***
	(0.015)	(0.011)
Year 1998	-0.0331	0.0192
	(0.022)	(0.019)
Year 2000	-0.0082	0.0095
	(0.021)	(0.018)
Year 2002	-0.0386	-0.0002
	(0.024)	(0.019)
Year 2004	-0.0087	0.0226
	(0.022)	(0.018)
Year 2006	0.0206	0.0346*
	(0.023)	(0.019)
Year 2008	0.0064	0.0561***
	(0.025)	(0.021)
Year 2010	0.0023	0.0475**
	(0.024)	(0.020)

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.

dy/dx for factor levels is the discrete change from the base level. Reference groups are unmarried, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health

insurance, no chronic health conditions, and year 1996. Arthritis has been omitted because of collinearity. Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7 consists of three parts; the coefficient estimates for the labor supply equation, the coefficient estimates for the health equation, and covariance-variance error terms estimates.

First, focusing on the results from 2SPS and FIML in the labor supply (Table 7), the coefficient estimates on health in the labor supply equation are positive and significant with the same accuracy ($p<0.01$) using both methods of estimation. However, the results from FIML have smaller magnitudes compared to 2SPS for both males (0.89, 2SPS; 0.68, FIML) and females (0.69, 2SPS; 0.60, FIML). The better the health, the more likely one is work.

We find that the effect of age will be positive, but decreasingly. As age increases, up to a certain age, then the effect becomes negative and increasingly so as age continues to increase. To see this, just evaluate 0.3328*age - 0.0039*age*age for females at various values, or take the derivative and solve for zero to find the point of inflection. The calculation shows that the point of inflection for females is age around 43, and for males the age is 36. The marginal effect of age, holding all other covariates at their mean is significant and negative for both males and females. The change in probability of labor supply associated with a one-year difference in age is around a 3% decrease ($p<0.01$) for males and a 2% decrease ($p<0.01$) for females. We also find that the effect of age on the labor supply is significant and negative for those older than 62 years for both males and females.

We use Table 8 to examine the change in probability of working arising from unit change in covariate, given all other covariates at their mean values. The presence of children younger than 18 has a negative insignificant effect on women's supply of labor. However, this effect is marginally significant and positive for males (0.03, $p<0.10$).

The change in probability of males' labor supply when marital status goes from 0 to 1 (the expected difference in probability of working) increases 4 percentage points or 0.04, and is significant ($p<0.01$). However, this change has a negative effect on females' labor supply, and is significant (-0.06, $p<0.01$). Thus, married males are more likely to work compared to married females.

The change in probability of males' labor supply increases as the level of education increases. The females' probability of working decreases for individuals with less than a high school education compared to those with high school diploma, and is significant (-0.06, $p<0.01$). The same finding is not significant for males. The probability of labor supply increases when the level of education changes from high school diploma to college/some college, and is marginally significant for males (0.03, $p<0.1$), and strongly significant for females (0.03, $p<0.01$). The probability of working increases by 7 percentage points when the level of education goes from high school diploma to more than college level for males, and is significant (0.07, $p<0.01$), and insignificant for females.

The change in probability of working is associated with a 6 percentage points decrease for males and 5 percentage points decrease for females, when occupation changes from white collar 1 workers [6] to the blue collar workers for males ($p<0.01$). The probability of working decreases by 3 percentage points when occupation changes from white collar 1 to white collar 2, is significant ($p<0.01$) for females, and insignificant for males.

The probability of females' labor supply decreases by 0.0013 percentage points as wealth increases ($p<0.01$). For males, the wealth effect is not statistically significant, but its coefficient sign is positive.

The probability of African American males' labor supply decreases by 4 percentage points ($p<0.05$), and it is insignificant for females.

[6] White collar 1 includes managerial specialty operation or technical support

The region of residency impact is not significant for males. For females, those who live in the Midwest and the Northeast are 4 and 6 percentage points more likely to work respectively, compared with those who live in the South.

Finally, the change in probability of working when employer provided health insurance goes from 0 to 1 increases by 18 and 30 percentage points respectively for males and females. In this section, we assume employer provided health insurance is exogenous with respect to labor supply. However, there is a possibility of endogeneity of health insurance to labor supply. To address this issue, we rerun the regressions for a group of married sample who have health insurance through their spouses.

Focusing on the results from the 2SPS and FIML methods in the health equation (Table 7), we find a positive and significant association between labor supply and health for males (0.87, 2SPS; 0.30, FIML) and females (0.02, 2SPS; 0.03, FIML). While the estimate from FIML is significant at the 5% level, the one from 2SPS method is insignificant for females. The estimate from both methods are significant at the 1% significant level for males.

Because our main focus in this study is the relationship between health and labor supply, we briefly discuss the estimates for other explanatory variables in the health equation using FIML method. The effect of age is negative and insignificant for both males and females. However, the positive and significant effect of age squared implies that the age effect becomes stronger as people age. The effect of marital status is insignificant for males, and strongly significant and positive for females. Married women are 12% more likely to be in better health ($p<0.01$). As it is common in the literature, education has a strong direct effect on health. Males and females with less than a high school education are 37% and 46% less likely to be in better health respectively compared with those with high school education ($p<0.01$). Males and females with college/some college education are 19% and 23%, and those with more than college education are 35% and 53% more likely to be in better health respectively. Males and females white collar 2 workers are 17% and 7% less likely to be in better health compared with white collar 1 workers ($p<0.01$). The same estimates for the blue collar workers are 17% and 25% ($p<0.01$). The effect of wealth is insignificant for males, and significant ($p<0.01$) for females. Both Hispanics and African Americans have poorer health outcomes compared with non-Hispanics Whites ($p<0.01$).

The census region is statistically significant for females but not for males. The poorer physical functionality and the higher number of chronic illnesses suggest poorer health outcomes for both males and females ($p<0.01$). Male smokers are 14% ($p<0.01$) and female smokers are 8% (0.10) less likely to have better health outcomes compared with nonsmokers. As it is expected, most chronic illnesses have a positive significant effect on health for both males and females. Moreover, those with health insurance are more likely to have better health status.

Finally, the following paragraph discusses possible source of endogeneity of health to labor supply using the results of FIML estimation method.

It has been argued both simultaneity and unobserved heterogeneity may result in the endogeneity of health to labor supply (Cai 2010). People may justify their non-employment status by reporting poor health or employment may have a negative effect on their health (simultaneity). In addition, there might be some other unobserved individuals fixed effect that have an adverse effect on both health and labor supply (unobserved heterogeneity). FIML estimation method enables us to conduct the true endogeneity test, and investigate the true source of endogeneity using a simultaneous equations model. Focusing on the results from the FIML estimation, the effect of health on the labor supply is statistically significant and positive. The reverse effect of the labor supply on health is positive and significant for both males and females. However, the estimated time-varying unobserved error terms ($\delta_{hL(v)}$) are negative and statistically significant. For both males and females, the positive effect of labor supply on health would lead an upward bias in the estimate of the effect of health on the labor supply. Meanwhile, the negative correlation between the time-varying error component of health and labor supply equations implies a downward bias in the estimate of the effect of health on labor supply. Overall, the ambiguous net effect suggesting that it is not possible to measure the direction of the bias caused by the endogeneity of health for both males and females. Our result supports the finding for females in the Cai (2010). The finding supports past literature that health should not be treated exogenous. However, we have not found any supportive evidence to determine the bias direction of health effect. In addition, the time-invariant error terms in both health and labor supply equations are very large. This result implies that using a panel data and controlling for unobserved heterogeneity improve the efficiency of estimation (Cai 2010)

2.8.2. 2SRI estimation method

As it has been argued by Blundell and Powell (2004), 2SRI method relies on the control function concept. To control for endogeneity of health with respect to labor supply, we use the residuals from the reduced form of the ordered probit estimation (health status equation) as covariates in the probit estimation (labor supply equation). Using simulation method, some studies showed that 2SRI has superior to common 2SPS in non-linear estimations (Basu and Coe 2017; Terza et al. 2008). As Basu and Coe (2017) have argued, we can use 2SRI estimator with varying forms of residuals to estimate non-linear models. We follow Vella (1993) to generate the generalized residuals.

Table 9 presents the coefficient estimates from the 2SRS method.

Table 9. Coefficient estimates 2SRI method.

	Male (N=17,391)	Female (N=26,798)
Labor supply Equation		
Health		
Fair	1.5400***	1.3029***
	(0.208)	(0.196)
Good	2.5770***	2.2471***
	(0.244)	(0.217)
Very good	3.3888***	2.9453***
	(0.301)	(0.253)
Excellent	4.3496***	3.4744***
	(0.388)	(0.323)
Generalized residuals	-0.0340***	-0.0267***
	(0.005)	(0.004)
Age	0.1184	0.3817***
	(0.225)	(0.130)
Age squared	-0.0023	-0.0044***
	(0.002)	(0.001)
Married	0.2694***	-0.2592***
	(0.103)	(0.074)
Child 0-18	0.1566	0.1365
	(0.422)	(0.320)
Married*Child 0-18	-0.0680	-0.3259
	(0.453)	(0.367)
> High school	0.1802	-0.2060**
	(0.141)	(0.104)
College/some college	-0.0316	0.0662
	(0.133)	(0.089)
< College	0.0872	-0.1400
	(0.150)	(0.114)
Age 62+	-0.5110**	-0.1990
	(0.251)	(0.193)
Less than high school*Age 62+	0.0015	0.1567
	(0.271)	(0.237)
College*Age 62+	0.1743	-0.0245
	(0.302)	(0.233)
More than college*Age 62+	0.2337	-0.0174
	(0.287)	(0.272)
White collar 2 occupation	0.0437	-0.0596
	(0.098)	(0.068)
Blue collar occupation	-0.2065**	-0.1425
	(0.089)	(0.091)
Wealth [a]	-0.0008	-0.0074
	(0.003)	(0.005)
Hispanic	0.3651***	0.2599***
	(0.113)	(0.086)
African American	-0.0620	-0.0030
	(0.091)	(0.068)
Midwest	0.0728	0.1873***
	(0.081)	(0.063)
Northeast	0.0643	0.2822***
	(0.098)	(0.077)
West	0.1139	-0.0279
	(0.092)	(0.073)

Employer provided HI	0.8380***	1.3533***
	(0.101)	(0.088)
Year 1998	0.1312	0.2079
	(0.160)	(0.137)
Year 2000	0.0397	0.0712
	(0.187)	(0.157)
Year 2002	-0.1703	-0.0093
	(0.202)	(0.164)
Year 2004	0.0985	0.1426
	(0.196)	(0.160)
Year 2006	0.1473	0.1253
	(0.215)	(0.170)
Year 2008	0.2055	0.2785
	(0.220)	(0.173)
Year 2010	0.0356	0.1482
	(0.203)	(0.171)
Constant	-0.3016	-8.4205**
	(6.017)	(3.336)
Health Equation		
Labor supply	0.4777***	0.3784***
	(0.182)	(0.130)
Generalized Residuals	-0.0814	-0.0897
	(0.095)	(0.075)
Physical functionality	-0.2092***	-0.2302***
	(0.033)	(0.029)
No. of chronic condition	-0.4018***	-0.4892***
	(0.047)	(0.045)
Age	-0.1141	-0.0855
	(0.077)	(0.072)
Age squared	0.0012	0.0010
	(0.001)	(0.001)
Married	-0.0018	0.1436***
	(0.060)	(0.050)
> High school	-0.3846***	-0.4327***
	(0.054)	(0.050)
College	0.2145***	0.2164***
	(0.048)	(0.041)
< College	0.4229***	0.5112***
	(0.056)	(0.053)
White collar 2 occupation	-0.1784***	-0.0557
	(0.054)	(0.045)
Blue collar occupation	-0.2407***	-0.2339***
	(0.053)	(0.059)
Wealth	0.0027	0.0045
	(0.002)	(0.004)
Hispanic	-0.4357***	-0.6463***
	(0.067)	(0.059)
African American	-0.1797***	-0.4200***
	(0.054)	(0.044)
Current smoking	-0.1374*	-0.0527
	(0.072)	(0.078)
Current drinking	-0.0739	-0.0721
	(0.138)	(0.305)
Lagged smoker	-0.0735	-0.0642
	(0.088)	(0.100)

Lagged heavy drinker	-0.0972	0.1039
	(0.170)	(0.328)
Lagged preventive behavior	0.0074	-0.0355
	(0.071)	(0.068)
Midwest	0.0519	0.0928**
	(0.045)	(0.042)
Northeast	0.0659	0.0547
	(0.054)	(0.051)
West	0.0024	0.0808*
	(0.052)	(0.048)
High blood pressure	-0.0044	-0.0155
	(0.065)	(0.066)
Diabetes	-0.1830**	-0.1230
	(0.079)	(0.084)
Cancer	-0.2593**	-0.0534
	(0.128)	(0.088)
Lung disease	-0.3583***	-0.2641**
	(0.120)	(0.113)
Heart disease	-0.2738***	-0.1268
	(0.077)	(0.088)
Stroke	0.0929	0.2884*
	(0.128)	(0.152)
Psychiatric problem	-0.1465	0.0088
	(0.099)	(0.083)
Health insurance	0.1559*	0.1241
	(0.089)	(0.082)
Cut-1	-5.8435***	-5.2451***
	(2.037)	(1.872)
Cut-2	-4.3372**	-3.5316*
	(2.036)	(1.871)
Cut-3	-2.7586	-1.9248
	(2.036)	(1.870)
Cut-4	-1.1299	-0.1291
	(2.035	(1.870)
$Ln(\delta_h)$	0.0280	0.0933***
	(0.022)	(0.017)
$Ln(\delta_l)$	0.5147***	0.4208***
	(0.038)	(0.029)
Log likelihood health Eq.	-20122	-29409
Log likelihood labor supply Eq.	-6994	-11281

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.

Reference groups are unmarried, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996. Arthritis has been omitted because of collinearity. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

The 2SRI marginal effects at means have been reported in Table 10.

Table 10. Marginal effects of labor supply equation using estimates from the 2SRI method.

Change in y given unit change in x	Male	Female
Health		
Fair	0.4581***	0.3579***
	(0.054)	(0.044)
Good	0.7954***	0.7075***
	(0.037)	(0.035)
Very good	0.9014***	0.8591***
	(0.035)	(0.034)
Excellent	0.9268***	0.9082***
	(0.032)	(0.033)
Age	-0.0237***	-0.0318***
	(0.004)	(0.004)
Married	0.0476**	-0.0703***
	(0.021)	(0.017)
Child 0-18	0.0156	-0.0259
	(0.023)	(0.048)
> High school	0.0287*	-0.0468**
	(0.015)	(0.022)
College/some college	0.0023	0.0148
	(0.015)	(0.016)
< College	0.0240	-0.0384
	(0.017)	(0.024)
Age 62+	-0.0750	-0.0510
	(0.046)	(0.048)
White collar 2 occupation	0.0061	-0.0152
	(0.014)	(0.017)
Blue collar occupation	-0.0344**	-0.0377
	(0.015)	(0.025)
Wealth [a]	-0.0001	-0.0019
	(0.001)	(0.001)
Hispanic	0.0486***	0.0607***
	(0.013)	(0.018)
African American	-0.0104	-0.0008
	(0.016)	(0.018)
Mideast	0.0121	0.0482***
	(0.013)	(0.016)
Northeast	0.0107	0.0694***
	(0.016)	(0.018)
West	0.0184	-0.0079
	(0.015)	(0.021)
Employer provided HI	0.1555***	0.3268***
	(0.021)	(0.020)
Year 1998	0.0212	0.0545
	(0.026)	(0.036)
Year 2000	0.0068	0.0199
	(0.032)	(0.043)
Year 2002	-0.0334	-0.0027
	(0.041)	(0.047)
Year 2004	0.0163	0.0385
	(0.032)	(0.043)
Year 2006	0.0236	0.0341
	(0.034)	(0.046)
Year 2008	0.0317	0.0707*

	(0.033)	(0.043)
Year 2010	0.0061	0.0399
	(0.035)	(0.045)

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.

dy/dx for factor levels is the discrete change from the base level. Reference groups are unmarried, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996. Arthritis has been omitted because of collinearity. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Here, we retain the observed health status while adding the residuals from the reduced form. Thus, instead of predicted value of health, we have health status as a categorical variable. As can been seen in Table 10, the change in the probability of working increases as health status becomes better.

Table 11 provides the predicted probability of labor supply, given 5 ranges of health status.

Table 11. Predicted conditional probability of labor supply using estimates from the 2SRI method

Health status	Predicted probability of labor supply (in percent)
Males	
Poor	7
Fair	52
Good	87
Very good	97
Excellent	99
Females	
Poor	9
Fair	48
Good	79
Very good	92
Excellent	96

The probabilities are conditional on the observed health status, and all other variables are at their mean values.

Focusing on males with poor health status, the probability of working is 9%, given that all predictors are set to their mean values. Overall, the predicted probability of working is smaller for females compared with males. As the health status becomes better, the probability of working increases.

Table 12. Predicted conditional probability of labor supply using estimates from the FIML method

Health status	Predicted probability of labor supply (in percent)
Males	
Poor	38
Fair	63
Good	81
Very good	91
Excellent	96
Females	
Poor	39
Fair	59
Good	74
Very good	84
Excellent	92

The probabilities are conditional on the estimated cut-off points for the observed health status.

The probability in Table 12 are calculated using the whole sample. That is why the results should not directly compare to Table 5. For instance, people with self-reported poor health status are more likely to also have other characteristics that adversely affect their labor supply.

2.9. Discussion

Using partial information maximum likelihood (2SPS and 2SRI) and full information maximum likelihood methods of estimation, we estimated the relationship between health and labor supply equations. We used three methods to check the possible endogeneity of health with respect to labor supply. The results of the true endogeneity test from FIML estimation confirmed the results of previous studies (Stern 1989; Cai 2010) that health should not be treated as exogenous. To address the endogeneity of health, a simultaneous equations model was used.

Using RAND longitudinal HRS data 1996-2010, we found a positive significant effect of health on labor supply for both males and females (0.6833, 0.5953; $p<0.01$ males and females respectively). The results also suggested a significant positive reverse effect of labor supply on health for both males and females (0.2981, $p<0.01$; 0.0305, $p<0.05$ males and females respectively).

Full Information Maximum Likelihood estimation method allows us to examine the potential bias in estimated effect of health. We can estimate the association between health and labor supply variables, together with the covariance-variance estimates. Our findings indicate that the potential bias arising from treating health as an exogenous variable could not be determined from the data, and the positive reverse effect may not result from justification. On one hand, the positive reverse effect of labor supply on health, and on the other hand, the significant negative correlation between the time-varying error components covariance (-1.0293, $p<0.01$; -0.6082, $p<0.01$ males and females respectively) in the health and labor supply equations leave us with no evidence on the direction of bias arising from endogeneity of health to labor supply. The result supports the finding for female but not for males in Cai (2010).

Our result also confirmed the finding in Cai (2010) that there were efficiency gains in using panel data. We found that the variances of the time-varying unobserved error components are large and highly significant in both equations ($Ln(\delta_h)$ and $Ln(\delta_l)$).

CHAPTER 3. EFFECT OF HEALTH INSURANCE ON LABOR SUPPLLY

3.1. Literature review

Some studies examined the role of health insurance in labor supply. Individuals with employer-provided insurance (EPHI) are less likely to reduce their supply of labor following a chronic health chronic condition. They may also have the incentives to remain employed and maintain health insurance. (Bradley et al. 2006; Bradley et al. 2012; Bradley et al. 2013; Zimmer 2015).

Bradley et al. (2007) found women who have health insurance through their spouse are more likely to leave labor market or if they remain, they tend to reduce their hours of work. Bradley et al. (2012) also found significant decline in employment of married men with ECHI. Bradley et al. (2013) found women who are newly diagnosed with breast cancer, and have ECHI are more likely to remain employed following a health shock. Zimmer (2015) studied the role of fringe benefit in the employment effects of health shocks. Using Medical Expenditure Panel Survey (MEPS), he found that full time employers are 4.4 percentage points less likely to remained employed, and 0.7 percentage points more likely to shift from full-time jobs to part-time, and 3.7 percentage points more likely to quit working.

Bradley et al. (2002) argued, employer provided health insurance may be endogenous to labor supply, and individuals who obtain health insurance through their employers are less willing to exit the labor market. To address the endogeneity of EPHI, one solution could be sampling a group of married, and estimate the effect of obtaining health insurance through spouse on labor supply. Then health insurance may not directly affect the individuals' decision about their labor supply. It is expected that a spouse who obtains health insurance from a source other than his/her employer be more willing to leave the labor market.

In the following sections, we examine the role of health insurance in the supply of labor. We answer the following question: what is the impact of health insurance provided by spouse's employer on a married individual? Since in the RAND HRS, we have information on the source of obtaining health insurance, we can simply include the spouse's health insurance to evaluate the effect of health insurance on the labor supply taking the endogeneity of health insurance to labor supply into account.

3.2. Descriptive statistics

Table 13 presents descriptive statistics in the pooled eight-wave sample for males and females.

Table 13. Variable means, a sample of married males and females.

	Male (N= 14,236)	Female (N= 19,132)
Health status	3.4085	3.5062
Physical functionality	-0.5326	-0.4626
Age	58.1634 (4.3472)	56.3637 (5.5058)
Child 0-18	0.1229	0.0957
> High school	0.1517	0.1200
College	0.2407	0.2684
< College	0.2927	0.2268
Age 62+	0.2591	0.1957
> High school & Age 62+	0.0487	0.0295
College & Age 62+	0.0541	0.0462
< College & Age 62+	0.0700	0.0364
White collar 2 occupation	0.2003	0.5363
Blue collar occupation	0.4466	0.1249
Wealth [a]	4.4835	4.5071
Hispanic	0.0959	0.0767
African American	0.1092	0.1059
Current smoker	0.6749	0.4964
Current heavy drinker	0.0514	0.0082
Lagged smoker	0.4875	0.3755
Lagged heavy drinker	0.0389	0.0066
Lagged preventive behavior	0.4006	0.4567
Midwest	0.2652	0.2631
Northeast	0.1499	0.1453
West	0.1906	0.1840

Chronic condition	1.3400 (1.2278)	1.3256 (1.2242)
High blood pressure	0.4360	0.3695
Diabetes	0.1536	0.1058
Cancer	0.0572	0.0853
Lung disease	0.0436	0.0487
Heart disease	0.1676	0.0963
Stroke	0.0353	0.0247
Psychiatric problem	0.0796	0.1388
Spouse's health insurance	0.1599	0.3708
Health insurance	0.9177	0.9131
Year 1998	0.1718	0.1630
Year 2000	0.1394	0.1387
Year 2002	0.1174	0.1218
Year 2004	0.1462	0.1441
Year 2006	0.1105	0.1180
Year 2008	0.0891	0.0944
Year 2010	0.0738	0.0747

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.

Means or sample percentages are reported with standard deviations of continuous variables in parentheses. Reference groups are, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996.

The HRS is a nationally representative sample of those aged 51 and older, but spouses are included in the data regardless of age. The sample of males is restricted to ages 22 to 64 and females to ages 23 to 64. Overall, the sample of males can be described as predominantly middle age (58.16), white (89 percent), have high school diploma (32 percent), have good health status or better (3.51), have health insurance (92 percent – 57 percent mostly covered by their own employers), and are employed (73 percent). Thirty-nine percent live in the South. The sample of females is also predominantly middle aged (mean age is 56.48), white (89 percent), have high school diploma (38 percent), have good health status or better (3.28), have health insurance (91 percent – 62 percent mostly not covered by their spouses' health insurance, and are employed (64 percent). Forty-one percent live in the South.

Table 14 tabulates labor supply status against self-reported health using the pooled sample of married 1996-2010.

Table 14. Labor supply status by self-reported health for a sample of married.

Employment status	Health status: Poor (1)	Fair (2)	Good (3)	Very good (4)	Excellent (5)	All
Females						
% non-employment	78.42	53.05	34.96	30.14	28.29	36.14
% employment	21.58	46.95	65.04	69.86	71.71	63.86
Observations	760	2,528	5,484	6,988	3,372	19,132
Males						
% non-employment	73.94	42.76	26.04	19.57	13.79	26.77
% employment	26.06	57.24	73.96	80.43	86.21	73.23
Observations	733	2,030	4,447	4,741	2,285	14,236

A positive relationship appears from a simple tabulation.

3.3. Results

Table 15 presents the coefficient estimates labor supply and health equations for a sample of married males and females. The negative effect of spousal employer health insurance on the probability of working is as one would expect. We also add the natural transformation of spouse's earnings to the model. Our result supports Bradley et al. (2002) finding. The effect of spouse's earnings on the probability of working is statistically significant and positive, which is surprising.

Table 15. Influence of spouse insurance on probability of working of a sample of married using FIML.

	Male (N=14,236)	Female (N=19,132)
Labor supply Equation		
Health	0.6657***	0.5718***
	(0.025)	(0.056)
Age	0.3923	0.4433***
	(0.365)	(0.120)
Age squared	-0.0046	-0.0049***
	(0.003)	(0.001)
> High school	0.1930**	-0.0506
	(0.086)	(0.103)
College/some college	0.0195	0.0827
	(0.079)	(0.074)
< College	0.0325	-0.1340
	(0.090)	(0.106)
White collar 2 occupation	0.0570	-0.0915
	(0.090)	(0.078)
Blue collar occupation	-0.1616*	-0.1407
	(0.088)	(0.110)
Wealth [a]	-0.0011	-0.0067
	(0.003)	(0.006)
Hispanic	0.3689***	0.2730***
	(0.111)	(0.103)
African American	-0.0193	0.4996***
	(0.089)	(0.088)
Midwest	0.0147	0.2037***
	(0.074)	(0.075)
Northeast	0.0327	0.3559***
	(0.089)	(0.092)
West	0.0186	-0.1452*
	(0.085)	(0.086)
Presence of child under 18	0.0651	-0.2697
	(0.112)	(0.183)
Spousal health insurance	-0.4731***	-0.7622***
	(0.102)	(0.091)
Spouse's earnings	0.0387***	0.0425***
	(0.007)	(0.009)
Constant	-3.5822	-7.3345**
	(5.155)	(3.181)
Health Equation		
Labor supply	0.2163***	-0.0049
	(0.050)	(0.027)
Physical functionality	-0.2407***	-0.3118***
	(0.033)	(0.041)
No. of Chronic conditions	-0.3733***	-0.4559***
	(0.051)	(0.052)
Age	-0.1689	-0.0707
	(0.306)	(0.067)
Age squared	0.0018	0.0007
	(0.002)	(0.001)
> High school	-0.3274***	-0.5134***
	(0.064)	(0.062)
College/some college	0.2080***	0.2188***
	(0.054)	(0.048)

< College	0.3726***	0.5170***
	(0.062)	(0.062)
White collar 2 occupation	-0.1649***	-0.0766
	(0.062)	(0.051)
Blue collar occupation	-0.1929***	-0.2575***
	(0.070)	(0.072)
Wealth	0.0031	0.0037
	(0.003)	(0.004)
Hispanic	-0.4584***	-0.6589***
	(0.085)	(0.073)
African Americans	-0.1556**	-0.4745***
	(0.062)	(0.057)
Current smoker	-0.1903***	-0.1303***
	(0.035)	(0.036)
Current heavy drinker	-0.0395	0.0684
	(0.113)	(0.327)
Midwest	0.0652	0.1475***
	(0.049)	(0.049)
Northeast	0.0797	0.1369**
	(0.060)	(0.061)
West	-0.0003	0.0962*
	(0.059)	(0.058)
High blood pressure	0.0523	-0.0189
	(0.061)	(0.076)
Diabetes	-0.1617**	-0.1986**
	(0.074)	(0.097)
Cancer	-0.2414**	-0.0831
	(0.120)	(0.105)
Lung disease	-0.3331***	-0.2918**
	(0.115)	(0.141)
Heart disease	-0.2480***	-0.2255**
	(0.071)	(0.104)
Stroke	-0.0744	0.1994
	(0.138)	(0.189)
Psychiatric problem	-0.1910*	-0.1269
	(0.101)	(0.096)
Health insurance	0.2437**	0.2369**
	(0.101)	(0.101)
Cut-1	-7.0515	-5.6393***
	(9.774)	(1.760)
Cut-2	-5.4943	-3.8934**
	(9.771)	(1.760)
Cut-3	-3.8660	-2.2377
	(9.768)	(1.758)
Cut-4	-2.1773	-0.3618
	(9.763)	(1.759)
Ln (δ_L)	0.3372***	0.4277
	(0.039)	(0.040)
Ln (δ_H)	0.0476*	0.1233***
	(0.025)	(0.022)
Time-invariant error term	-0.6022***	-0.3386***
	(0.072)	(0.047)
Time-variant error term	-0.9119***	-0.5467***
	(0.039)	(0.060)
Log likelihood	-21936	-29012

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.

Reference groups are no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996. Arthritis has been omitted because of collinearity. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 16 presents the probability of labor supply derivatives with respect to the independent variables, evaluated at the mean values.

Table 16. Marginal effects for a group of married males and females.

Change in y given unit change in x	Male	Female
Age	0.3269	0.4018***
	(0.200)	(0.117)
Age squared	-0.0040**	-0.0045***
	(0.002)	(0.001)
Child 0-18	0.0760	-0.2689
	(0.132)	(0.183)
> High school	-0.0292	-0.3432***
	(0.083)	(0.089)
College/some college	0.1845**	0.2072***
	(0.076)	(0.068)
< College	0.3276***	0.1611*
	(0.087)	(0.089)
White collar 2 occupation	-0.0617	-0.1349*
	(0.086)	(0.074)
Blue collar occupation	-0.3388***	-0.2871***
	(0.080)	(0.104)
Wealth [a]	0.0011	-0.0046
	(0.002)	(0.004)
Hispanic	0.0745	-0.1035
	(0.101)	(0.097)
African American	-0.1436	0.2277***
	(0.088)	(0.081)
Midwest	0.0679	0.2873***
	(0.072)	(0.070)
Northeast	0.1002	0.4329***
	(0.087)	(0.086)
West	0.0215	-0.0900
	(0.083)	(0.082)
Spouse's health insurance	-0.5527***	-0.7601***
	(0.120)	(0.091)
Spouse's earning [b]	0.0452***	0.0424***
	(0.008)	(0.009)
Physical functionality	-0.1872***	-0.1778***
	(0.027)	(0.030)
No. of chronic condition	-0.2903***	-0.2599***
	(0.038)	(0.033)
Current smoker	-0.1480***	-0.0743***
	(0.028)	(0.023)
Current heavy drinker	-0.0307	0.0390
	(0.088)	(0.186)
High blood pressure	0.0406	-0.0108
	(0.048)	(0.043)
Diabetes	-0.1257**	-0.1132**
	(0.057)	(0.055)
Cancer	-0.1877**	-0.0474
	(0.093)	(0.060)
Lung disease	-0.2590***	-0.1664**
	(0.090)	(0.082)
Heart disease	-0.1928***	-0.1286**
	(0.056)	(0.060)
Stroke	-0.0579	0.1137
	(0.108)	(0.107)

Psychiatric problem	-0.1485*	-0.0723
	(0.078)	(0.057)
Health insurance	0.1895**	0.1351**
	(0.077)	(0.060)
Constant	-4.1847	-7.3142**
	(5.900)	(3.158)
Log likelihood	-21936	-29012

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Spouse's earning is measured by its natural logarithm.

dy/dx for factor levels is the discrete change from the base level. Reference groups are unmarried, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/Other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996. Arthritis has been omitted because of collinearity. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 17 presents the marginal effect of spouse's health insurance on the probability of working conditional on five ranges of latent health using the estimated parameters from the FIML.

Table 17. Marginal effect of spouse's health insurance on probability of labor supply at five ranges of health status.

Health status	Marginal effect of spouse's health insurance	
	Males	Females
Poor	-7.00e-06	-0.0064***
	(0.00004)	0.0016
Fair	-0.0022***	-0.0538***
	(0.0012)	0.0100
Good	-0.0672***	-0.1749***
	(0.0233)	0.0252
Very good	-0.2696***	-0.3027***
	(0.0646)	(0.0364)
Excellent	-0.2429***	-0.2996***
	(0.0479)	0.03448

AME, Average Marginal Effect. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Focusing on the results of Average Marginal Effect (AME), when adjusted to the sample distribution of other model covariates, the average decrease in the probability of working for males is about -6 percentage points for those who have health insurance through their spouses and have good health. For females, the average decrease in the probability of working is about -17 percentage points for those who are insured through their spouses compared with those who don not have health insurance through their spouses. The effects are bigger for females than males in terms of magnitude. In other words, holding the health status equal, females are more willing to leave the labor market when they are insured by their husbands' health insurance.

3.4. Discussion

We found evidence for the significant role of health insurance in the decision to work among married women and men (-0.5527, $p<0.01$; -0.7601, $p<0.01$ males and females respectively). Our result is consistent with the finding in Bradley et al. (2002). Also, our finding suggested that women with spouse's health insurance are more likely to exit the labor market compared with men. The change in probability of working with respect to the spouse's health insurance increases, given the five ranges for observed health status.

CHAPTER 4. CANCER AND LABOR SUPPLY

4.1. Background

Cancer has a major impact on society in the United States, and is the second highest cause of death. According to National Cancer Institute (NCI) the number of people surviving after a cancer diagnosis reached around 14.5 million in 2014 and is expected to rise to around 19 million by 2024. The five-year survival rates of those diagnosed with cancer have increased by 20 percentage points among whites and 24 percentage points among African American. However, among patients aged 50 to 64 years, improvements in survival have been more pronounced than among older than 65 Siegel et al. (2017).

Advances in the treatment of cancer and as a result an increase in cancer survival rate have induced researchers to examine the effect that cancer has on survivors' employment. Among all type of cancer, breast cancer received the most attention in the literature. Bradley et al. (2005) found that breast cancer survivors are less likely to work within the six months after being diagnosed with cancer. Ganz et al. (1996) studied quality of life of survivors' breast cancer two and three years after treatment. They found that women who have breast cancer and have survived, continue to work and they perform their usual work even after being diagnosed with breast cancer. They found that 65% of those breast cancer survivors who were working were functioning at a high level. Satariano et al. (1996) examined the likelihood of returning to work after breast cancer, and they found that 71 percent of women with breast cancer returned to their work within three months after diagnosis.

Overall, cancer has a negative impact on cancer survivors' employment compared with non-cancer control group (Bradley et al. 2002; Bradley et al. 2012). However, the literatures on short- and long-run impacts have shown different results in terms of significance and size of impact. Bradley et al. (2002) examined the effect of breast cancer on women's labor supply as well as hours of work. They found that the survivors of breast cancer are 10 percentage points less likely to be employed compared with a non-breast cancer group. Surprisingly, breast cancer survivors work around three more hours per week than women who do not have cancer. Short et al. (2005) studied the cancer survivors' employment 1-year to 5-year from the time of diagnosis to follow-up. They found that 13% of cancer survivors have quit working after diagnosis with cancer within 4 years of diagnosis, more than half quit working after one year.

In this chapter, we examine the effect of cancer on the probability of working for samples of males and females. We specifically investigate the short- and long-term impacts of cancer on labor supply by taking advantage of panel data and including the number of years after diagnosis with cancer in our analyses. Taking advantage of panel data improves the efficiency of estimates by increasing the degrees of freedom and reducing the collinearity among explanatory variables. We also examine the effect of cancer on hours of work to test the finding in Bradley et al. that breast cancer survivors who remain in labor market work more than non-cancer group. We use a married sample to control for true effect of health insurance on the probability of working. We also test the hypothesis that those cancer survivors who are married and insured by their spouses are less likely to work than married non-cancer group and without spouse's health insurance.

4.2. Data

We use data from the 1996-2010 HRS and RAND HRS data. The HRS is a longitudinal data that allows us to study the interested outcome by looking at past and post-event, and is one of the most valid and rich data set. The criteria we use to sample the data is similar to the chapter 2. However, we merge the information on the date of diagnosis with cancer from 1996-2010 HRS to the RAND HRS version to capture the short- and long-term impacts of cancer on the probability of working and hours of work. After excluding respondents who are older than 65 and missing information, there are n=7,551 total individuals with information on N=30,020 observations. The data are unbalanced because the entry times to the study for different respondents are not the same. In the HRS dataset, only one member of the household must be above 51; Therefore, we have spouses in the survey even if their age is below 51.

To constitute our non-cancer group, we select respondents who replied “no” to the question, “Has a doctor ever told you that you have cancer or a malignant tumor of any kind?” Subsequently, those who replied “yes” are our cancer group. One of limitation is that the HRS does not publicly provide the information on type of cancer.

Table 18 presents the descriptive statistics for a group of married males and females. The prevalence estimate of cancer is 4.3 percent for males (n=562), and 5.2 for females (n=855). The average time since diagnosis is 5.07 years (SD=5.6 range 1-38) for males, and 7.76 years (SD=7.72 range 1-40) for females.

Table 18. Descriptive information.

	Males			Females		
	Cancer	Non-cancer	Total	Cancer	Non-cancer	Total
	(N=562)	(N=12,273)	(N=12,835)	(N=855)	(N=16,330)	(N=17,185)
Labor supply	67.79%***	75.60%	75.26%	55.91%***	64.72%	64.28%
Age	59.59***	58.04	58.11	56.97***	56.25	56.29
Years since diagnosis	5.07			7.76		
> High school	10.32%***	15.64%	15.40%	11.70%***	13.66%	13.56%
High school	28.47%*	32.75%	46.86%	37.78%*	37.79%	37.79%
College/some	30.78%	22.86%	23.21%	25.50%	26.39%	26.35%
< College	30.43%***	28.75%	28.82%	25.03%***	22.16%	22.30%
White/other	89.32%***	89.51%	89.50%	89.94%***	88.59%	88.65%
Black	10.68%***	10.49%	10.50%	10.06%***	11.41%	11.35%
Hispanics	3.20%***	10.60%	10.28%	5.15%***	9.65%	9.43%
Wealth	5.54***	4.42	4.46	3.87***	4.33	4.31
Hbp [c]	45.73%***	41.15%	41.35%	43.04%***	35.96%	36.32%
Diabetes	14.95%***	13.09%	13.17%	12.40%***	9.77%	9.90%
Lung	2.85%***	3.50%	3.47%	6.55%***	3.96%	4.09%
Heart	20.11%***	13.85%	14.13%	9.94%***	8.66%	8.72%
Stroke	2.85%***	2.87%	2.87%	3.86%***	2.06%	2.15%
Psychiatric	12.10%***	7.39%	7.60%	16.49%***	14.06%	14.18%
Smoking	70.82%***	66.24%	66.44%	52.75%***	47.66%	47.91%
Drinking	2.67%***	5.17%	5.06%	0.70%***	0.85%	0.84%
White collar1	37.37%***	34.11%	34.25%	36.96%***	33.00%	33.19%
White collar2	26.16%	19.61%	19.90%	54.97%	53.93%	53.98%
Blue collar	36.48%***	46.28%	45.85%	8.07%***	13.08%	12.83%
Spouse's HI	19.93%*	15.29%	15.49%	37.19%*	36.18%	36.23%
Hours work	40.94%***	43.45%	43.34%	35.93%***	36.20%	36.77%
Spouse's earning[b]	5.32***	5.87	5.85	5.85***	6.01	6.00

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Spouse's earning is measured by its natural logarithm.
[c] Hpb is High blood pressure.

Means or sample percentages are reported with standard deviations of continuous variables in parentheses. *Significantly different from sample of women without breast cancer at p<0.10, **p<0.05, ***p<0.01.

Overall, this sample can be described as predominantly white, and middle-aged (mean age is 58 and 56 for males and females, respectively), and as having a high school education or better. Males and females with cancer are significantly different from the control group in their age. Far fewer people with cancer work (68% versus 76%, 56% versus 64% for males and females, respectively), but those work and have cancer, work nearly the same hours per week (41 versus 43 for males and 36 for females in both groups).

4.3. Empirical approach

We follow Bradley et al. (2002) approach. However, we use a probit model for a panel instead of a cross-sectional analysis. The outcomes of interest are employment status (E) and weekly hours of work (H). Aside from the key variable, we control for other explanatory variables that would affect the decision to work including demographic and financial characteristics. We control for age, level of education (less than high school, high school, college/some college, more than college), race/ethnicity (African-Americans, white/other, Hispanic, and non-Hispanic), spouse's health insurance, spouse's earning, type of occupation (white collar 1, white collar 2, blue collar), presence of children younger than 18, and wealth (the value of respondent's housing equity plus non-housing equity). The main explanatory variable is the incidence of cancer. The impact of cancer can be estimated in the model by using either a binary variable (yes=1 no=0) or as an ordinal variable that indicates the years of diagnosis with cancer as a categorical variable, or the number of years after diagnosis with cancer as a continuous variable. We include it as a categorical variable because as it has been discussed in Bradley et al. (2002), the impact of cancer on employment is not linear. The employment equation is as follows:

$$E_{it}^{*} = f(CA_{it}, X_{it}, SHI_{it}, SE_{it}, CD_{it}) \tag{17}$$

We define employment as a binary variable that equals one if the respondent replies "yes" to the question "are you currently working for pay", otherwise zero. Cancer (CA_{it}) is an ordinal variable ranging from 0 to three (0=no cancer, 1= two or fewer years prior to interview, 2=three or between three and five years, 3= five or more years) for an individual i at time t. Exogenous variables (X_{it}), availability of spouse's health insurance, the natural log of spouse's earning (SE_{it}), and other chronic health diseases (CD_{it}). The employment equation is estimated using a probit model for panel data. Thus, the probit estimates are reported as marginal effects of the independent variables.

We also estimate the average weekly hours worked for those employed. It is common to use Heckman's sample-selection models Heckman (1979) to estimate such models when the dependent variable is censored. Hours censored for those who do not work ($H_{it} = 0 \; if \; E_{it} = 0$), where H_{it} is defined as average weekly hours work for individual i at year t, and as follows for those who are employed and reported positive hours,

$$H_{it} = \beta_0 + \beta_1 CA_{it} + X_{it}\beta_2 + \beta_3 SHI_{it} + \beta_4 SE_{it} + \beta_5 CD_{it} + \varepsilon_i \quad if \; E_i = 1 \tag{18}$$

Following Bradley et al. (2002), we assume a linear functional form and also the same variables that affect employment status will affect the hours worked equation.

4.4. Results

Table 19 reports males' and females' probabilities of working for a sample of married.

Table 19. Probit coefficient estimates, a sample of married respondents

	Male (N=12,936)	Female (N=17,669)
Years since diagnosis with cancer (Yr)		
$0 < Yr \leq 2$	-0.2204	-0.4267***
	(0.180)	(0.142)
$3 \leq Yr < 5$	-0.1496	-0.2502
	(0.204)	(0.177)
$Yr \geq 5$	-0.4757**	-0.1374
	(0.213)	(0.120)
Age	-0.1798***	-0.1218***
	(0.010)	(0.005)
Child 0-18	0.1632	-0.5229***
	(0.109)	(0.078)
> High school	-0.0884	-0.5229***
	(0.126)	(0.106)
College/ some	0.2145*	0.1885**
	(0.110)	(0.084)
< College	0.4315***	0.2647***
	(0.120)	(0.101)
African American	-0.1297	0.0407
	(0.135)	(0.104)
Hispanic	0.2715**	-0.2796**
	(0.138)	(0.118)
Wealth [a]	0.0044	-0.0113***
	(0.003)	(0.003)
High blood pressure	-0.1393*	-0.0883*
	(0.071)	(0.053)
Diabetes	-0.4960***	-0.2936***
	(0.095)	(0.082)
Lung disease	-0.6643***	-0.5380***
	(0.179)	(0.122)
Heart disease	-0.5320***	-0.3446***
	(0.103)	(0.091)
Stroke	-0.8845***	-0.7297***
	(0.226)	(0.171)
Psychiatric problems	-0.8832***	-0.6608***
	(0.136)	(0.074)
Smoking	-0.1654*	-0.2215***
	(0.087)	(0.065)
Drinking	0.1496	0.0612
	(0.121)	(0.194)
White collar occupation 2	-0.0986	-0.1034
	(0.115)	(0.078)
Blue collar	-0.5349***	-0.2887**
	(0.105)	(0.113)
Spouse's health insurance	-0.6068***	-0.7817***
	(0.085)	(0.044)
Spouse's earning [b]	0.0625***	0.0562***
	(0.006)	(0.004)
Constant	12.2287***	8.1066***
	(0.604)	(0.300)
δ_μ	1.1778***	1.0472***

	(0.075)	(0.052)
Number of respondents	3,391	4,160
Log likelihood	-4965	-7838

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Spouse's earning measured by its natural logarithm.

Robust standard errors in parentheses. Reference groups are no children under 18 residing with respondent, high school diploma, white collar 1 occupation, non-Hispanic, white/other, no spouse's health insurance, no chronic health conditions, no smoking, and no drinking.
*** p<0.01, ** p<0.05, * p<0.1.

Table 20. Marginal effects of probit coefficient estimates.

Change in y given unit change in x	Males	Females
Cancer years since diagnosis (Yr)		
$0 < Yr \leq 2$	-0.0331	-0.0830***
	(0.028)	(0.028)
$3 \leq Yr < 5$	-0.0222	-0.0481
	(0.031)	(0.035)
$Yr \geq 5$	-0.0745**	-0.0262
	(0.036)	(0.023)
Age	-0.0261***	-0.0230***
	(0.001)	(0.001)
Child 0-18	0.0237	-0.0986***
	(0.016)	(0.015)
> High school	-0.0138	-0.1030***
	(0.020)	(0.021)
College/ some	0.0317**	0.0351**
	(0.016)	(0.016)
< College	0.0612***	0.0489***
	(0.017)	(0.018)
African American	-0.0192	0.0076
	(0.020)	(0.019)
Hispanic	0.0377**	-0.0538**
	(0.018)	(0.023)
Wealth [a]	0.0006	-0.0021***
	(0.000)	(0.001)
High blood pressure	-0.0203*	-0.0167*
	(0.010)	(0.010)
Diabetes	-0.0767***	-0.0565***
	(0.015)	(0.016)
Lung disease	-0.1068***	-0.1052***
	(0.031)	(0.025)
Heart disease	-0.0824***	-0.0665***
	(0.017)	(0.018)
Stroke	-0.1463***	-0.1441***
	(0.041)	(0.035)
Psychiatric problems	-0.1444***	-0.1289***
	(0.024)	(0.015)
Smoking	-0.0240*	-0.0417***
	(0.013)	(0.012)
Drinking	0.0217	0.0115
	(0.018)	(0.037)
White collar 2 occupation	-0.0131	-0.0193
	(0.015)	(0.014)
Blue collar occupation	-0.0775***	-0.0549**
	(0.015)	(0.022)
Spouse's health insurance	-0.0946***	-0.1493***
	(0.014)	(0.009)
Spouse's earning [b]	0.0091***	0.0106***
	(0.001)	(0.001)

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Spouse's earning measured by its natural logarithm.

dy/dx for factor levels is the discrete change from the base level. Reference groups are no children under 18 residing with respondent, high school diploma, white collar 1 occupation, non-Hispanic, white/other, no spouse's health insurance, no chronic health conditions, no smoking, and no drinking. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

The marginal effects of probability of working show that holding all covariates at their mean, the effect of cancer is negative (-0.07) statistically significant (p<0.01) for males three or more than three and less than five years since diagnosis. The effect of cancer is negative for those two or less years since diagnosis for both males (-0.03) and females (-0.08). However, the effect is only significant for females (0.01). Also, the estimates for those more recently diagnosed is larger (-0.03 verses -0.02, males; -0.08 verses -0.04, females). This finding supports the finding in Bradley et al. (2002). The coefficient estimates are negative but not statistically significantly different from zero for those who are diagnosed between three and five or three years for both male and females. However, the effects are negative. Those males with diagnosis five or more than five years prior to interview are 7 percentage points less likely to work compared with a non-cancer group (p<0.05).

For females, those diagnosed with cancer more than five years are less likely to work. However, the effect is not statistically significant. As the findings suggest, for females the short run impact of cancer is larger and significant while for males the same impact is larger and significant as the number of years since diagnosis increase. Overall, the effect of cancer is larger for female than males.

Other chronic health conditions have negative, statistically significant effect on the probability of working, and The effects are larger for males than females. Those who smoke and blue collar workers are less likely to work in both samples. African American are less likely to work. However, the results are not significant for both males and females. Hispanic males are 4% more likely to work while Hispanic females are 5% less likely to work (p<0.05). Our finding supports the finding in Bradley et al. that surprisingly, the spouse's earning has a significant positive effect on the probability of working. However, the size of effect is negligible for both males and females (0.01, p<0.01). Borrowing from Bradley et al. (2002), this finding may reflect "assortative mating or complementarities in the consumption of time of older men and women" (p. 1319).

The spouse's health insurance effect is negative and significant for both males (-0.09, p<0.01) and females (-0.15, p<0.01), and the effect is larger for females than males (9% verses 15%). In another specification, we also test the hypothesis that married with cancer and spouse's health insurance are less likely to work by including an interaction term. The marginal effect of probit estimates is reported in Appendix D. Here, we measure the incidence of cancer as a binary variable.

Table 21. Influence of spouse's health insurance on the probability of cancer group verses non-cancer group, a sample of married.

	Males (N=12,936)	Females (N=17,669)
Spouse's health insurance	-0.6232***	-0.7719***
	(0.088)	(0.050)
Cancer	-0.3394**	-0.2034
	(0.141)	(0.125)
Cancer*Spouse's health insurance	-0.7273**	-1.0939***
	(0.292)	(0.163)
Age	-0.1801***	-0.1215***
	(0.010)	(0.006)
Child 0-18	0.1632	-0.5229***
	(0.109)	(0.097)
> High school	-0.0910	-0.5227***
	(0.126)	(0.108)
College/some	0.2141*	0.1878**
	(0.110)	(0.084)
< College	0.4290***	0.2657**
	(0.120)	(0.104)
African American	-0.1285	0.0387
	(0.135)	(0.105)
Hispanic	0.2734**	-0.2812**
	(0.138)	(0.113)
Wealth [a]	0.0044	-0.0113***
	(0.003)	(0.003)
High blood pressure	-0.1399**	-0.0884
	(0.071)	(0.060)
Diabetes	-0.4972***	-0.2919***
	(0.095)	(0.093)
Lung disease	-0.6637***	-0.5422***
	(0.179)	(0.143)

Heart disease	-0.5406***	-0.3423***
	(0.103)	(0.102)
Stroke	-0.8875***	-0.7362***
	(0.224)	(0.252)
Psychiatric problem	-0.8816***	-0.6595***
	(0.135)	(0.089)
Smoking	-0.1675*	-0.2198***
	(0.087)	(0.065)
Drinking	0.1476	0.0644
	(0.121)	(0.216)
White collar 1	-0.0995	-0.1043
	(0.115)	(0.083)
Blue collar	-0.5336***	-0.2880**
	(0.105)	(0.121)
Spouse's earning [b]	0.0626***	0.0561***
	(0.006)	(0.005)
Constant	12.2500***	8.0890***
	(0.605)	(0.384)
δ_μ	1.1773***	1.0458***
	(0.075)	(0.057)
Log likelihood	-4966	-7840

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Spouse's earning measured by its natural logarithm.

Reference groups are no children under 18 residing with respondent, high school diploma, white collar 1 occupation, non-Hispanic, white/other, no spouse's health insurance, no chronic health conditions, no smoking, and no drinking. Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

As can be seen in Table 21, those males and females with cancer and have health insurance through their spouses are less likely to work, and the effect is larger for sample of female. The same finding in Bradley et al. (2002) is insignificant.

Table 22 provides the results from the Heckman sample selection model for hours of work. Since our data have a panel structure, the regular Heckman command in Stata does not work. Then, I program to first estimate the selection equation via xtprobit, and get estimates on the mills ratio. I then use the mills ratio as an explanatory variable in the wage equation where only the truncated dependent variable is considered, and estimate this equation for selection variable (hours<.) equal one. However, since we conduct the two-stages analysis manually, we need to correct the standard errors in the second stage by bootstrapping both regressions (xtprobit and xtreg) simultaneously (Woodridge 2002), and account for the the inter-equation correlation between the error terms. The number of replication is set at 50.

Table 22 presents the bootstrapped results for the hours of work.

Table 22. Coefficient estimates with bootstrapped standard errors using Hackman sample estimation for weekly hours worked, conditional on working

	Male	Female
Cancer years since diagnosis (Yr)		
0 <Yr≤ 2	-2.7625*	-1.4554
	(1.4537)	(1.7882)
3 ≤Yr< 5	-0.6420	-4.6972**
	(1.7167)	(2.0451)
Yr≥ 5	-2.2475	-2.7032
	(2.5314)	(1.7195)
Age	-0.6066***	-0.1780
	(0.0764)	(0.1491)
Child 0-18	0.6186	0.3796
	(0.9129)	(1.4338)
> High school	-2.4687**	0.5563
	(1.1576)	(1.3865)
College/some	1.1876	0.8310
	(1.2759)	(1.0143)
<College	1.1948	1.5039
	(1.2645)	(1.0649)
African-American	-1.7686	0.1505
	(1.2660)	(1.0985)
Hispanic	-0.9630	0.2934
	(1.2862)	(1.2235)
Wealth [a]	0.0303	-0.0404
	(0.0347)	(0.0609)
High blood pressure	0.5433	0.2334
	(0.9187)	(1.0263)
Diabetes	0.1657	0.3039
	(1.1569)	(1.3411)
Lung disease	-0.8741	1.1828
	(2.2299)	(2.2221)
Heart disease	1.4681	0.2254
	(1.6152)	(1.4226)
Stroke	1.2982	0.8192
	(3.5048)	(3.5847)
Psychiatric problems	0.8501	0.9212
	(2.3402)	(1.2878)
Smoking	-1.7305**	1.0297
	(0.8758)	(0.6652)
Drinking	0.6909	-3.8166
	(1.0294)	(2.4068)
White collar 2	0.3463	-1.7637*
	(0.9436)	(0.9880)
Blue collar	0.9663	0.5531
	(0.8397)	(1.4310)
Spouse’s health insurance	-1.4196*	-1.9357***
	(0.8302)	(0.5973)
Spouse’s earning [b]	-0.0151	-0.0050
	(0.0958)	(0.0678)
Constant	78.8828***	45.9686***
	(4.3948)	(8.0207)

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Spouse's earning measured by its natural logarithm.

Reference groups are no children under 18 residing with respondent, high school diploma, white collar 1 occupation, non-Hispanic, white/other, no spouse's health insurance, no chronic health conditions, no smoking, and no drinking. Bootstrapped standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

The result suggests that employed men who are diagnosed with cancer for two years or less, work 2.76 hours less a week (p<.10) than employed men without cancer. For females, employed women with cancer diagnosis for two years or fewer, work 1.46 hours less a week, but the result is not statistically significant. Our finding is in contrary to the finding in Bradley et al. study in terms of the direction of coefficient. They found that breast cancer employed women work more. For those cancer survivors who are employed and diagnosed with cancer three years or between three and five years, the direction of effect on hours of work is negative. However, the result for males is insignificant, and for females, the result indicates that they work 4.70 hours a week less than the non-cancer employed group (p<0.05). Also, the results are not significant for the five years or more diagnosis with cancer. However, the effect is negative. We also find that employed men with spouse's health insurance, work 1.42 hours less (p<0.10) than employed men without spouse's health insurance. For females, employed women who have health insurance through their husbands, work 1.94 hours less compared with those women without spousal health insurance, and the result is highly significant (p<0.01).

4.5. Discussion

We examined the short- and long-term impacts of cancer on the labor market outcomes for a sample of married people. We found that the probability of working for the women cancer survivors is 8 percentage points (p<0.01) less in the short-run (two years or fewer since diagnosis) and the probability of working for men cancer survivors is in association with a 7 percentage points (p<0.01) reduction in three or between three to five years since diagnosis. Employed men in the years immidiately following diagnosis, work 2.76 hours (p<0.10) less a week than other employed men. Employed women following three to five years since diagnosis, work 4.70 hours ($p<0.05$) less per week.

CHAPTER 5. SURVIVAL ANALYSIS; EFFECT OF CANCER ON DURATION OF NON-EMPLOYMENT

5.1. Background

Individuals with poor health are more likely to be observed among the unemployed (Arrow 1996). It has been argued that poor health increases the risk of unemployment so that the duration of unemployment spells significantly is longer for those individuals have impaired health (Stewart 2001). García-Gómez et al. (2010) studied the role of health in employment in order to estimate the effect of health on the hazard of becoming employed and non-employed. They found that health status affects employment in terms of entering into and exiting out of labor market, and the effect is bigger for males compared with females.

In this chapter, we are only intereted in testing whether incidence of cancer is an obstacle to re-employment. Then an appropriate sample could be one consisting of non-employed people who are looking for job. Thus, the duration of non-employment becomes our interested dependent variable and health shock, here incidence of cancer, is the risk factor. We expect that the cancer survivors group experience longer duration of non-employment following the health shock.

There are two approaches to estimate the duration of non-employment; non-parametric and parametric estimates. We first illustrate the non-parametric estimates including Hazard ratio, Nelson-Aalen cumulative hazard estimate, Kaplan-Meier survival estimate by groups, Life time table. Then, we analyze the model using Cox parametric estimation method.

5.2. Data

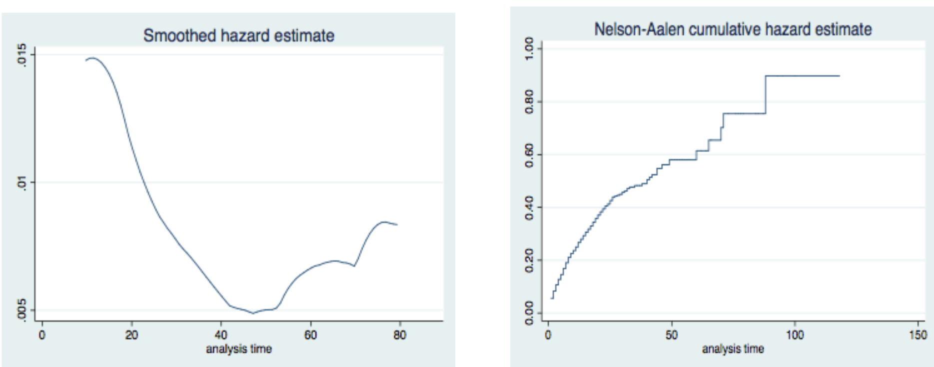
We use the 1996-2012 HRS data. To conduct a survival analysis, we define our failure event to equal 1 once the individual transmit from non-employment status to employment status, and otherwise zero. Then, we are interested in the duration of this transition. The duration is measured in months, and we use two dates in the HRS as stop previous job and start current job. The respondents replied to the question "In what year and month did you stop working at your previous job?" and "In what year and month did you start working at your current job?". After merging all the HRS files, and eliminating the missing values on the key variables, and limiting the age to below 65, and those who are not self-employed, there are 7,503 observations.

5.3. Empirical approach

5.3.1. Non-parametric approach

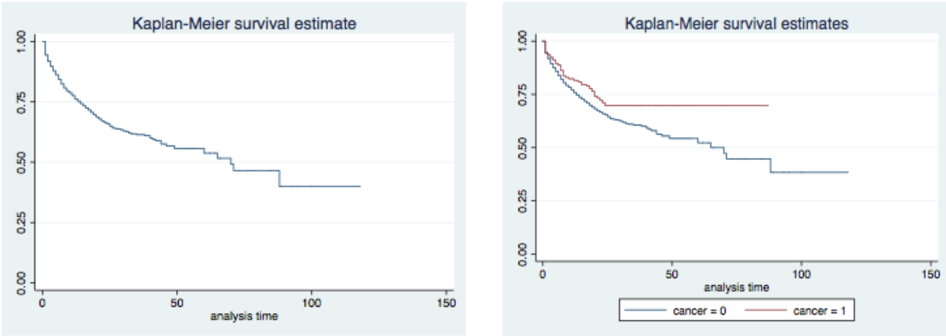
First, we focus on the non-parametric approach, which includes Hazard-Ratio (HR), Nelson-Aalen cumulative hazard estimate, Kaplan-Meier survival estimate, and Life Time table. In survival analysis, the hazard ratio (HR) is defined as the number of failure events per unit time divided by the number at risk. In our study, the HR is defined as the number of individuals who experience the event 1 (a transition from the non-employment status to the employment status) divided by the total number of individuals at risk. Figure 2 shows the hazard-ratio.

Figure 2. Hazard Ratio



The Hazard ratio graph represents, as the time goes by, people are less likely to start a job. The cumulative hazard ratio graph sums up the hazard ratio over time. Next, we use Kaplan-Meier survival estimate to show the survival estimate by cancer group verses non-cancer group. Figure 3 represents the result form the KM survival estimate.

Figure 3. Kaplan-Meier survival estimates by cancer.



The Kaplan-Meier (KM) graph always starts at 1.00 and it declines as time goes by because at the beginning the full sample experiences unemployment and at the end of the study more than 30 percent of the sample does not experience the event or find a job. Using the graph, we can say for instance at the month 50 more than 50 percent of the sample have not found a job. Using the KM graph, we can compare two groups of cancer and non-cancer. For instance, those who had cancer experiencing much higher survival rate than those who did not have cancer. It implies that cancer group were experiencing a longer non-employment period. The same analysis can be done for those who receive unemployment compensation during their unemployment.

Lastly, a life table summarizes survival data in terms of the number of events and the proportion surviving at each event time point. Table 23 represents some part of the life time table.

Table 23. Life-Time Table

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[99% Conf. Int.]	
1	6543	312	235	0.9523	0.0026	0.9450	0.9587
2	5996	149	184	0.9287	0.0032	0.9199	0.9365
3	5663	117	218	0.9095	0.0036	0.8997	0.9183
4	5328	93	212	0.8936	0.0039	0.8831	0.9032
5	5023	77	221	0.8799	0.0041	0.8688	0.8901
6	4725	93	202	0.8626	0.0044	0.8507	0.8735
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
100	8	0	1	0.4580	0.0521	0.3209	0.5847
101	7	0	1	0.4580	0.0521	0.3209	0.5847
105	6	0	1	0.4580	0.0521	0.3209	0.5847
107	5	0	1	0.4580	0.0521	0.3209	0.5847
118	4	0	1	0.4580	0.0521	0.3209	0.5847
122	3	0	1	0.4580	0.0521	0.3209	0.5847
151	2	0	1	0.4580	0.0521	0.3209	0.5847
194	1	0	1	0.4580	0.0521	0.3209	0.5847

5.3.2. Parametric approach

This section presents the results of the Cox's proportional hazards model following Arrow (1996). As it is argued by Arrow (1996), Cox estimation does not model the dependent variable, here, non-employment duration, but it models a function of it, which is the natural logarithm of the hazard rate function instead of the classical linear regression. Assuming that the Cox's proportional hazards model is,

$$\lambda(t|X_1, \dots, X_p) = \lambda_0(t)e^{(\beta_1 X_1 + \cdots + B_p X_P)} \tag{19}$$

Where $\lambda_0(t)$ is the baseline hazard and not to be specified for the parameter estimation. by taking logarithm, we will have a linear function,

$$\log\left[\lambda(t|X_1, \dots, X_p)\right] = \log\left[\lambda_0(t)\right] + \beta_1 X_1 + \cdots + \beta_p X_p) \tag{20}$$

We estimate the $\beta_1, \dots, \beta_p$ coefficients, where are the incidence of cancer, demographic variables, the years of schooling, wealth, and receipt of unemployment compensation.

We conduct the equality survival test (Log-Rank test) to test the equality of survival function of cancer group and non-cancer group.

Table 24. Equality test of survival time for cancer verses non-cancer groups

Cancer	Event observed	Event expected
No	1325	1274.96
Yes	103	153.04
Total	1428	1428.00

The null hypothesis is, the cancer and non-cancer groups have the same survival. The expected number of subjects at each time point in each is adjusted for the number of subjects at risk in the groups at each event time. The log-rank test determines if the observed number of events in each group is significantly different from the expected number. The formal test is based on a chi-squared statistic. When the log-rank statistic is large, it is evidence for a different in the survival times between the groups. The log-rank statistic has a chi-squared distribution with one degree of freedom, and the p-value is calculated using the chi-square distribution. The log-rank test for difference in survival gives a p-value of p=0.00, indicating that the cancer group do differ significantly in survival.

The following table represents the summary statistics per subject.

Table 25. Survival time summary statistics per subject.

		……………………....Per subject…………………....			
Category	Total	Mean	Min	Median	Max
No. of subject	6552				
No. of records	7503	1.1451	1	1	4
(First) entry time		0.0081	0	0	22
(Final) exit time		12.9245	1	11	194
Subject with gap	0				
Time on gap if gap	0				
Time at risk	84,628	12.9164	1	11	194
Failures	1,428	0.02179	0	0	1

The Mean of non-employment spell per subject was 1.1451 months. The Min of non-employment spell per subject was 1 month. The Max of non-employment spell per subject was 4 months. The number of subjects that experienced the failure event was 1,428. The total time at risk was 84,628. The incidence rate of finding a job was .0169. The total number of subjects at risk was 6,552. The 25 percentile of survival time was 18 months, and 50 percentiles was 88 months. The Cox model has been used to estimate the duration of being non-employed measured in months.

Table 26 represents the coefficients of the Cox model and corresponding risk ratios.

Table 26. Coefficient estimates in Cox proportional hazards model for non-employment duration.

Variable	Parameter estimate	Risk Ratio
Age	-0.0745***	0.9282***
	(0.0045)	(0.0042)
Female	-0.1838***	0.8321***
	(0.0585)	(0.0487)
African-American	-0.3352***	0.7152***
	(0.0826)	(0.0591)
Hispanic	-0.0511	0.9502
	(0.0961)	(0.0913)
Years of Schooling	0.0797***	1.0829***
	(0.0111)	(0.0120)
Unemployment compensation	0.2742***	1.3155***
	(0.0519)	(0.0682)
Cancer	-0.2121**	0.8089**
	(0.1129)	(.0913)
Log Likelihood	-9763.6933	-9763.6933

Standards errors shown in parentheses.
*p<0.1, **p<0.05, ***p<0.01.

A negative coefficient has a positive impact on the duration of non-employment, which suggesting a longer duration of non-working and a lower probability of finding a job.
One observes that the risk of transition from the non-employment status to employment-status is lower for females, African-American, Hispanics; with a relative risk of 0.8321, 0.7152, and 0.9502, respectively ($p<0.05$). The finding, as we expected, implies that females, African-Americans, and Hispanics are less at risk than males, whites, and non-Hispanics. This finding supports the finding in Hoffman (1991) that females, African-Americans, and older persons are expected to have a lower probability of escaping non-employment because of discrimination against them. There is also a significant effect due to age. As people age, the risk of escaping non-employment decrease with the relative hazard of 0.9282 ($p<0.01$). one would expect a longer non-employment spell for those who receive unemployment compensation. However, our result does not support this expectation. Higher educated people at a lower higher risk of escaping the non-employment status as was expected (relative risk of 1.0829), and the effect is highly significant ($p<0.01$).

Lastly, the incidence of cancer has a significant negative effect on the non-employment. Those who are diagnosed with cancer experience a longer non-employment spell compared with a non-cancer group (relative risk of 0.8089).

5.4. Discussion

This chapter examined initial analyses on the variables influencing the probability that non-employed individuals become employed. It has been showed that cancer is positively associate with the risk of non-employment, and those with cancer are at the higher risk of being non-employed. In addition, our results are consistent with Arrow (1996), and what one would expect from gender, and race inequalities. This chapter indicated a preliminary study on the impact of cancer on non-employment spell, and our results come from an initial analysis. Thus, there are several limitations to this study. We need to improve the results by controlling more variables, and conducting separate analyses for males and females. We also need to find a more relevant dataset other than the HRS for the survival analysis. However, it could be considered as a good starting point.

CHAPTER 6. CONCLUSION

This dissertation investigates the relationship between health and labor market outcomes. Health is often considered as an important variable in the individual's labor supply. Various methods of estimation have been used to examine this relationship. This study is divided into two main sections; the effect of self-reported health status on the labor supply, and the impact of cancer that would be considered as a health shock (less subjective measure of health) on the labor market outcomes. Advances in treatments for cancer have made cancer survivors more likely to stay in the labor market, and as a result make an inquiry into the impact of cancer on cancer survivors' labor supply more relevant. We have done all analyses separately for males and females due the fact that factors affect men's and women's labor supplies differently.

Few studies have examined the effect of health on the labor supply, considering health endogenous. Using using a simultaneous equation model, and a very rich panel dataset, we estimate the effect of self-reported health on the labor supply as we treat health endogenous to labor supply. The endogeneity of health comes from two different sources; the simultaneity and the unobserved heterogeneity. Employing a Full Information Maximum Likelihood (FIML) model allows for testing the justification hypothesis as well as the correlation between two equations error terms. Also, since we can estimate the variance-covariance matrix using FIML method, we can truly conduct the endogeneity test of health. The result showed that health is endogenous to labor supply, and applying the FIML method is more efficient compared with two-stage methods of estimation including two stage prediction substitute (2SPS) and two stage residual inclusion (2SRI). The finding from all methods of estimation indicated that there is a highly significant and positive effect of health on labor supply as we expected for both males and females. the coefficient estimates on health in the labor supply are significant and positive for both males and females at 1% significance level (0.68 males; 0.60

females). The reverse effect is also highly significant and positive for both males and females.

The effect of labor market on health associates with 0.30 ($p<0.05$) and 0.03 ($p<0.05$) percentage points for males and females, respectively. Although the endogeneity test of health suggests that if health treated exogenous then the estimated effect of health is likely to be biased, our results do not show the direction of bias. Because the reverse effect is positive, and simultaneously, the correlation between two equations' time-varying error terms are significantly negative, it looks impossible to estimate the bias in the effect of health on labor supply. This finding agrees with the finding in Cai (2010) for females, but not for males. The effect of many other factors including demographic and socio-economic variables, employer provider health insurance, individuals' lifestyle (smoking, drinking, use of preventive care) have been estimated.

The results using estimates from the FIML method indicated that predicted conditional probability of labor supply for males and females with poor health outcome are 38% and 39%, respectively. And the predicted conditional probability of labor supply decreases as the health outcome improves. We also, looked at the predicted conditional probability of labor supply using the 2SRI method. The results from 2SRI are smaller compared with the results from FIML. The predicted probability of working for men and female workers are 7% and 9% verses 38% and 39%. Because in the FIML we use the estimated values of health cut-off points while in the 2SRI we use the observed value of health status to estimate the predicted probability of working. That is why we shouldn't directly compare the results of theses two methods of estimation. It is highly likely that people with poor health outcome also have other characteristics that adversely affect their labor supply such as low income and a lower level of education.

Assuming employer provided health insurance (EPHI) is exogenous with respect to labor supply, the results suggest a significant and positive association between EPHI and labor supply. The increase in probability of working for males is associated with an 18 percentage points and for females is associated with a 30 percentage points at the 1% significance level for those who are insured by their employers compared with ones who have no EPHI. Thus, our results are in agreement with so called employment lock—workers with EPHI maintain higher labor supply to secure their health insurance coverage.

In another examination we try to account for the potential endogeneity of employer health insurance and consider the impact of spousal health insurance on the labor supply of two samples of married males and females. The findings suggest that the availability of spousal health insurance reduces the labor supply of both married men and women. We found evidence that spousal health insurance results in a 5.5 percentage points reduction in married men's probability of labor supply, and a 7.6 percent points reduction in married women's labor supply. Thus, as we expected, the labor supply reduction will tend to be smaller absolute value for males compared to females.

This finding has important implications for the health care reform in the United States. Cebi and Wang (2013) found a 2.6 percentage points reduction in women's labor supply. However, they did not use a simultanous equation model to determine health and labor coefficients jointly. Also, Bradley et al. (2002) suggests a 11 percentage points reduction in women's pobability of working. As Cebi and Wang (2013) argued, the results from models that do not account for unobserved effcets leads to a larger size of impact in terms of absolute value. The findings have important policy implications for health care reform and consequetly on the labor market outcomes. As the results suggest as we shift from employer-provided health insurance to spousal helath insurance, and disconnect the direct link between health insurance and employment, the labor supply decreases. Thus, an increase in the avalaivility of public health insurance may result in a more reduction in the probablity of working.

The short- and long-term impacts of cancer on the labor supply and worked hours of the married samples of males and females were also examined in this study. We specified the years since diagnosis with cancer to control for the short- and long-term effcets. We found that in the short-term, the women's probability of working is in association with an 8 percentage points reduction ($p<0.01$), but not significant short-term results for males. However, the long-term impact is significant and large for males, but not for females. The men's probability of working is in association with a 7 percentage points ($p<0.01$) reduction in three or between three to five years since diagnosis.

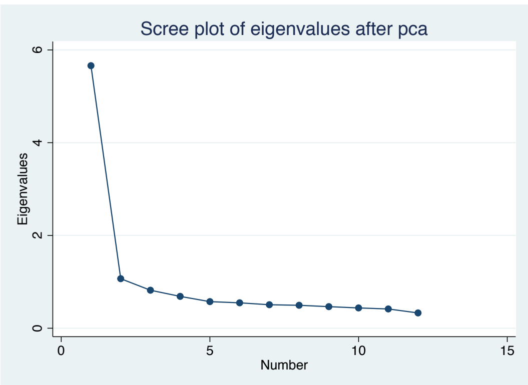
Employed men in the years immidiately following diagnosis, work 2.76 hours ($p<0.10$) less a week than other employed men. Emploed women following three to five years since diagnosis, work 4.70 hours ($p<0.05$) per week. Our findings are not in agreement with Bradely et al. (2002). They found employed breast cancer survivors are 3.39 hours more likely to work per week. We may not compare these two results because our focus is on cancer not breast cancer. We found strong evidence that cancer survivors are more likely to experience a longer non-employment spell than non-cancer group (relative risk of 0.81). Whether a cancer diagnosis impacts the labor supply and the duration of non-employment is important for public policy. The reduction in the labor supply following a cancer diagnosis cuase cancer survivors face a greater economic burden. Policy makers may find the results of short-, medium- and long-term impacts of cancer on the labor supply of cancer survivors beneficial to formulate policy that decrease this economic burden.

APPENDIX A- PRINCIPAL COMPONENT ANALYSIS (PCA) OF THE 10 ITEMS IN ADL AND IADL

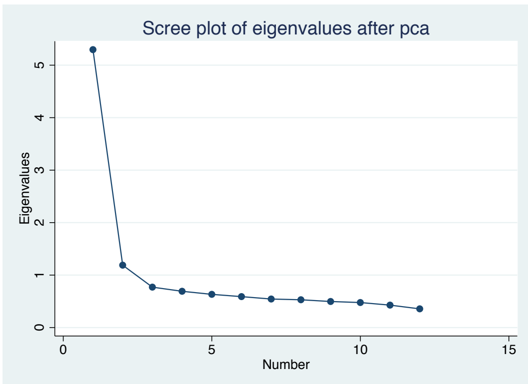
Appendix A1. Principal Component Analysis (PCA) Eigenvalue, performed in each wave (1996-2010)

	Eigenvalue (variance)							
Year	1996	1998	2000	2002	2004	2006	2008	2010
Component								
Walking across a room	5.66025	5.78789	5.73961	5.82251	5.56427	5.72015	5.49555	5.29585
	(0.4717)	(0.4823)	(0.4783)	(0.4852)	(0.4637)	(0.4767)	(0.458)	(0.4413)
Dressing	1.06756	1.12411	1.14688	1.09508	1.1513	1.12217	1.16583	1.18875
	(0.089)	(0.0937)	(0.0956)	(0.0913)	(0.0959)	(0.0935)	(0.0972)	(0.0991)
Bathing	0.818751	0.748852	0.733086	0.724871	0.753955	0.735852	0.737471	0.770059
	(0.0682)	(0.0624)	(0.0611)	(0.0604)	(0.0628)	(0.0613)	(0.0615)	(0.0642)
Eating	0.686153	0.650309	0.675996	0.658129	0.700532	0.631545	0.696228	0.691452
	(0.0572)	(0.0542)	(0.0563)	(0.0548)	(0.0584)	(0.0526)	(0.058)	(0.0576)
Getting in or out of bed	0.572949	0.571447	0.604383	0.57789	0.589778	0.584547	0.624655	0.633027
	(0.0477)	(0.0476)	(0.0504)	(0.0482)	(0.0491)	(0.0487)	(0.0521)	(0.0528)
Using the toilet	0.547387	0.544148	0.573028	0.55241	0.56378	0.567606	0.584468	0.589503
	(0.0456)	(0.0453)	(0.0478)	(0.046)	(0.047)	(0.0473)	(0.0487)	(0.0491)
Using a map	0.505971	0.514858	0.526786	0.519021	0.508275	0.534119	0.540303	0.543915
	(0.0422)	(0.0429)	(0.0439)	(0.0433)	(0.0424)	(0.0445)	(0.045)	(0.0453)
Using a telephone	0.494523	0.48375	0.466226	0.476393	0.496276	0.466546	0.487404	0.529677
	(0.0412)	(0.0403)	(0.0389)	(0.0397)	(0.0414)	(0.0389)	(0.0406)	(0.0441)
Handling money	0.464976	0.452724	0.435666	0.466552	0.471329	0.459764	0.463774	0.49669
	(0.0387)	(0.0377)	(0.0363)	(0.0389)	(0.0393)	(0.0383)	(0.0386)	(0.0414)
Taking medication	0.437114	0.40196	0.418921	0.41416	0.453359	0.427899	0.452739	0.476966
	(0.0364)	(0.0335)	(0.0349)	(0.0345)	(0.0378)	(0.0357)	(0.0377)	(0.0397)
Shopping	0.415094	0.398916	0.389981	0.403761	0.424095	0.416347	0.430286	0.427516
	(0.0346)	(0.0332)	(0.0325)	(0.0336)	(0.0353)	(0.0347)	(0.0359)	(0.0356)
Preparing meals	0.329269	0.321033	0.289432	0.289217	0.323056	0.333453	0.321292	0.3566
	(0.0274)	(0.0268)	(0.0241)	(0.0241)	(0.0269)	(0.0278)	(0.0268)	(0.0297)

Appendix A2. Scree plot of eigenvalues after PCA, Year 1996



Appendix A3. Scree plot of eigenvalues after PCA, Year 1996



APPENDIX B- MAXIMUM SIMULATED LIKELIHOOD (MSL)

Cai (2010) following Hyslop (1999) defines the log-likelihood function for the unknown vector parameter θ, given the random sample of observations as,

$$L_N(\theta) = \sum_{i=1}^{N} \ln\big(L(\theta; x_i)\big) \tag{B.1}$$

Then the maximum likelihood estimator for θ will be,

$$\hat{\theta}_{MSL} = \arg max_{\theta} \sum_{i=1}^{N} ln\tilde{L}(\theta; x_i, \xi_i) \tag{B.2}$$

where $\tilde{L}(\theta; x_i, \xi_i)$ is is an unbiased simulator for $L(\theta; x_i)$ and is defined as ,

$$(1/\mathrm{R})\sum_{r=1}^{R} \tilde{L}(\theta; x_i, \xi_{ri}) \tag{B.3}$$

where R is the number of simulation replication of the model.

APPENDIX C- FIRST STAGE ESTIMATIONS 2SPS METHOD

Appendix C1. The first stage health equation.

Health	Male	Female
Physical functionality	-0.2317***	-0.2479***
	(0.030)	(0.028)
Age	-0.0689	-0.0712
	(0.091)	(0.095)
Age squared	0.0006	0.0007
	(0.001)	(0.001)
Married	0.0256	0.1085**
	(0.064)	(0.051)
Child 0-18	0.2381	-0.0689
	(0.272)	(0.235)
Married*Child 0-18	-0.1863	0.0376
	(0.286)	(0.257)
> High school	-0.4273***	-0.4629***
	(0.084)	(0.075)
College/some college	0.2303***	0.2409***
	(0.069)	(0.060)
< College	0.4552***	0.5498***
	(0.078)	(0.070)
Age 62+	-0.0150	0.0101
	(0.144)	(0.141)
> High school *Age 62+	0.1188	-0.0155
	(0.181)	(0.182)
College/some college*Age 62+	0.0278	-0.0209
	(0.176)	(0.170)
< College *age 62+	0.0320	-0.0916
	(0.167)	(0.183)
White collar 2 occupation	-0.1728***	-0.0697
	(0.055)	(0.045)
Blue collar occupation	-0.2562***	-0.2543***
	(0.051)	(0.060)
Wealth [a]	0.0029	0.0040
	(0.002)	(0.004)
Hispanics	-0.4305***	-0.6519***
	(0.069)	(0.060)
African American	-0.1994***	-0.4222***
	(0.054)	(0.045)
Current smoker	-0.2270**	-0.1310
	(0.093)	(0.092)
Current heavy drinker	-0.0760	-0.0888
	(0.142)	(0.315)
Lagged smoker	0.0364	0.0271
	(0.122)	(0.120)
Lagged heavy drinker	-0.1008	0.0699
	(0.175)	(0.335)
Lagged preventive behavior	-0.0112	-0.0242
	(0.099)	(0.114)
Midwest	0.0601	0.1109***
	(0.045)	(0.042)
Northeast	0.0730	0.0793

	(0.055)	(0.051)
west	0.0128	0.0774
	(0.053)	(0.049)
No. of chronic conditions	-0.4228***	-0.5109***
	(0.046)	(0.045)
High blood pressure	0.0114	-0.0016
	(0.065)	(0.067)
Diabetes	-0.1862**	-0.1324
	(0.080)	(0.085)
Cancer	-0.2719**	-0.0598
	(0.131)	(0.090)
Lung diseases	-0.3807***	-0.2925**
	(0.124)	(0.115)
Heart diseases	-0.3011***	-0.1414
	(0.077)	(0.089)
Stroke	0.0537	0.2722*
	(0.130)	(0.155)
Psychiatric problems	-0.1686*	-0.0241
	(0.099)	(0.083)
Employer provided health insurance	0.2187***	0.0440
	(0.057)	(0.052)
Health insurance	0.0498	0.1291
	(0.101)	(0.087)
Year 1998	-0.3189**	-0.3470***
	(0.128)	(0.134)
Year 2000	-0.1653	-0.1277
	(0.136)	(0.125)
Year 2002	-0.1539	-0.1723
	(0.156)	(0.159)
Year 2004	-0.2279*	-0.2127*
	(0.126)	(0.118)
Year 2006	-0.1450	-0.1222
	(0.163)	(0.168)
Year 2008	-0.2749*	-0.2482*
	(0.156)	(0.145)
Year 2010	-0.1697	-0.1024
	(0.171)	(0.176)
Cut-1	-5.5456**	-5.6591**
	(2.337)	(2.431)
Cut-2	-4.0381*	-3.9467
	(2.337)	(2.431)
Cut-3	-2.4506	-2.3329
	(2.337)	(2.430)
Cut-4	-0.8107	-0.5247
	(2.337)	(2.430)
ln (δ_{μ})	0.0468**	0.1064***
	(0.021)	(0.017)
Observations	17,391	26,798
Log likelihood	-20109	-29372

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.

Reference groups are no children under 18 residing with respondent, high school diploma, white collar 1 occupation, non-Hispanic, white/other, no spouse's health insurance, no chronic health conditions, no smoking, and no drinking. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Appendix C2. The first stage health equation.

	Male	Female
Physical functionality	-0.3685***	-0.2866***
	(0.032)	(0.021)
Age	0.0784	0.3822***
	(0.160)	(0.073)
Age squared	-0.0018	-0.0044***
	(0.001)	(0.001)
Married	0.2575***	-0.2572***
	(0.082)	(0.056)
Child 0-18	0.2759	0.1210
	(0.300)	(0.208)
Married*Child 0-18	-0.1378	-0.3014
	(0.322)	(0.229)
> High school	0.0395	-0.3063***
	(0.116)	(0.086)
College/some college	0.1231	0.1846***
	(0.101)	(0.072)
< College	0.3262***	0.0654
	(0.115)	(0.087)
Age 62+	-0.5260***	-0.2038*
	(0.144)	(0.108)
> High school * Age 62+	0.0336	0.2019
	(0.176)	(0.148)
College * Age 62+	0.1792	-0.0148
	(0.168)	(0.132)
< College * Age 62+	0.2453	-0.0313
	(0.165)	(0.145)
White collar 2 occupation	-0.0514	-0.1363**
	(0.092)	(0.062)
Blue collar occupation	-0.3633***	-0.2755***
	(0.085)	(0.087)
Wealth [a]	0.0005	-0.0060***
	(0.003)	(0.002)
Hispanics	0.0692	-0.0547
	(0.112)	(0.090)
African American	-0.2248**	-0.1147*
	(0.092)	(0.069)
Current smoker	-0.2721**	-0.0920
	(0.109)	(0.082)
Current drinker	-0.0132	-0.2044
	(0.136)	(0.228)
Lagged smoker	0.0507	-0.0697
	(0.118)	(0.092)
Lagged drinker	-0.0187	-0.2860
	(0.156)	(0.257)
Lagged preventive behaviors	-0.0617	0.0576
	(0.096)	(0.085)
Midwest	0.1228	0.2460***
	(0.077)	(0.060)
Northeast	0.0846	0.3216***
	(0.091)	(0.070)
West	0.1062	-0.0013

	(0.087)	(0.067)
No. chronic conditions	-0.2718***	-0.1929***
	(0.064)	(0.049)
High blood pressure	0.2071**	0.1299*
	(0.092)	(0.073)
Diabetes	-0.0432	-0.1483*
	(0.106)	(0.090)
Cancer	-0.1457	-0.0554
	(0.138)	(0.095)
Lung	-0.1813	-0.3087***
	(0.151)	(0.113)
Hear disease	-0.2889***	-0.1783*
	(0.104)	(0.093)
Stroke	-0.4521***	-0.2380
	(0.167)	(0.155)
Psychiatric problem	-0.3414***	-0.3803***
	(0.128)	(0.087)
Employer provided HI	0.9424***	1.4748***
	(0.069)	(0.055)
Health insurance	-0.1632	-0.3202***
	(0.103)	(0.074)
Year 1998	-0.0318	0.0514
	(0.129)	(0.103)
Year 2000	-0.0527	0.0455
	(0.135)	(0.096)
Year 2002	-0.2090	-0.0647
	(0.150)	(0.119)
Year 2004	-0.0163	0.1056
	(0.127)	(0.092)
Year 2006	0.1308	0.0918
	(0.164)	(0.124)
Year 2008	0.0663	0.2696**
	(0.154)	(0.110)
Year 2010	-0.0107	0.1389
	(0.160)	(0.128)
Constant	2.8261	-6.4168***
	(4.413)	(1.924)
Log likelihood	-6892	-11121

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.

Reference groups are unmarried, no children under 18 residing with respondent, high school diploma, age under 62, white collar 1 occupation, non-Hispanic, white/other, South, no employer provided health insurance, no health insurance, no chronic health conditions, and year 1996. Arthritis has been omitted because of collinearity. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

APPENDIX D- MARGINAL EFFECTS OF PROBIT COEFFICIENT ESTIMATES

Appendix D1. Marginal effect of probit coefficient estimates after including an interaction term between the incident of cancer and spouse's health insurance.

Change in y given unit change in x	Males	Females
Incidence of cancer	-0.0462**	-0.0473**
	(0.021)	(0.022)
Spouse's health insurance	-0.0953***	-0.1492***
	(0.014)	(0.009)
Age	-0.0261***	-0.0229***
	(0.001)	(0.001)
Child18_	0.0237	-0.0986***
	(0.016)	(0.018)
> High school	-0.0142	-0.1030***
	(0.020)	(0.021)
College/ some	0.0316**	0.0350**
	(0.016)	(0.016)
< college	0.0608***	0.0491***
	(0.017)	(0.019)
African American	-0.0190	0.0073
	(0.020)	(0.020)
Hispanic	0.0380**	-0.0541**
	(0.018)	(0.022)
Wealth	0.0006	-0.0021***
	(0.000)	(0.001)
High blood pressure	-0.0204*	-0.0167
	(0.010)	(0.011)
Diabetes	-0.0769***	-0.0562***
	(0.015)	(0.018)
Lung disease	-0.1067***	-0.1061***
	(0.031)	(0.029)
Heart disease	-0.0839***	-0.0661***
	(0.017)	(0.020)
Stroke	-0.1469***	-0.1455***
	(0.041)	(0.051)
Psychiatric problem	-0.1441***	-0.1287***
	(0.024)	(0.018)
Smoking	-0.0243*	-0.0415***
	(0.013)	(0.012)
Drinking	0.0214	0.0121
	(0.018)	(0.041)
White collar 1	-0.0132	-0.0195
	(0.015)	(0.016)
Blue collar	-0.0774***	-0.0548**
	(0.015)	(0.023)
Spouse's earning [b]	0.0091***	0.0106***
	(0.001)	(0.001)

[a] Wealth is computed as total of all assets excluding the secondary residence, and divided by 100,000.
[b] Spouse's earning is measured by its natural logarithm.

dy/dx for factor levels is the discrete change from the base level. Reference groups are no children under 18 residing with respondent, high school diploma, white collar 1 occupation, non-Hispanic, white/other,

no spouse's health insurance, no chronic health conditions, no smoking, and no drinking. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

REFRENCES

Arrow, J. O. (1996). Estimating the influence of health as a risk factor on unemployment: A survival analysis of employment duration for workers surveyed in the German socio-economic panel (1984-1990). Social Science and Medicine , 42, 1651-1659.

Basu, A., & Coe, N. (2017). 2SLS vs 2SRI: appropriate methods for rare outcomes and/ or rare exposures. Health Economics , 26, 1087.

Bradley, C. J., Bednarek, H. L., & Neumark, D. (2002). Breast cancer and women's labor supply. Health Service Research , 37, 1309-1328.

Bradley, C. J., Neumark, D., & Barkowski, S. (2013). Does employer-provided health insurance constrain labor supply adjustments to health shocks? New evidence on women diagnosed with breast cancer. Health Economics , 32, 833 – 849.

Bradley, C. J., Neumark, D., & Motika, M. (2012). The effects of health shocks on employment and health insurance: the role of employer-provided health insurance. International Health Care Finance Economics , 12, 253-267.

Bradley, C. J., Neumark, D., Bednarek, H. L., & Schenk, M. (2005). Short-term effects of breast cancer on labor market attachment: Results from a longitudinal study. Health Economics , 24, 137-160.

Bradley, C. J., Neumark, D., Luo, Z., & Bednarek, H. (2007). Employment-contingent health insurance, illness, and labor supply of women: Evidence from married women with breast cancer. Health Economics , 16, 719-737.

Cai, B., Small, D. S., & Have, T. R. (2011). Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. Statistics in Medicine , 30, 1809-1824.

Cai, L. (2010). The relationship between health and labour force participation: evidence from a panel data simultaneous equation model. Labour Economics , 17, 77-90.

Cai, L., & Kalb, G. (2006). Health status and labour force participation: evidence from Australia. Health Economics , 15, 241-261.

Cebi, M., & Wang, C. (2013). Employer-provided health insurance and labor supply of married women. Eastern Economic , 39, 493-510.

Christensen, B. J., & Kallestrup-Lamb, M. (2012). The impact of health changes on labor supply: evidence from merged data on individual objective medical diagnosis codes and early retirement behavior. Health Economics , 21, 56-100.

Ganz, P. A., Coscarelli, A., Fred, C., Kahn, B., Polinsky, M., & Peterson, L. (1996). Breast cancer survivors: psychosocial concerns and quality of life. Breast Cancer Research and Treatment , 38, 183-99.

García-Gómez, P., Jones, A. M., & Rice, N. (2010). Health effects on labor market exits and entries. Labour Economics , 17, 62-76.

García-Gómez, P., Kippersluis, H. V., O'Donnell, O., & Doorslaer, E. v. (2013). Long-term and spillover effects of health shocks on employment and income. Human Resources , 48, 873-909.

Ginneken, J. K., & Groenewold, G. (2012). A single vs. multi-item self-rated health status measure: a 21-country study. The Open Public Health , 5, 1-9.

Grossman, M. (1972). On the concept of health capital and the demand for health. 80, 223-255.

Hajivassilio, V., & Ruud, P. (1994). Classical estimation methods for LDV models using simulation (Vol. 4). Amesterdam: Elsevier Scince Publishers.

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica , 46, 1251-1271.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica , 47, 153-161.

Hoffman, E. P. (1991). Estimation of length of job search by survival analysis. Eastern Economic , 17, 393-401.

Jean, M. M., & Burkhauser, V. R. (1990). Disentangling the effetc of arthritis on earnings: a simultaneous estimate of wage rates and hours worked. Applied Economics , 22, 1291-1309.

Lee, L. F. (1982). Health and wage: a simultaneous equation model with multiple discrete indicators. International Economic Review , 23, 199-221.

Levy, H., & Meltzer, D. (2008). The impact of health insurance on health. Annual Review of Public Health , 29, 399-409.

Roodman, D. (2011). Estimating fully observed recursive mixed-process models with cmp. Stata Journal , 11, 159-206.

Satariano, W. A., & DeLarenze, G. N. (1996). The likelihood of returning to work after breast cancer. Public Health , 111, 236-241.

Short, P. F., Vasey, J. J., & Tunceli, K. (2005). Employment pathways in a large cohort of adult cancer survivors. Cancer , 103, 1292–1301.

Siegel, R. L., Miller, K. D., & Ahmedin, J. (2017). Cancer statistics, 2017. CA: A Cancer Journal for Clinicians , 67, 7-30.

Smith, R., & Blundell, R. (1986). An exogeneity test for a simultaneous equation tobit model with an application to labor supply. Econometrica , 54, 679-685.

Stern, S. (1989). Measuring the effect of disabilit on labor force participation. Human Resources , 24, 361.

Stewart, J. M. (2001). The impact of health status on the duration of unemployment spells and implications for studies of the impact of unemployment on health status. Health Economics , 20, 781-796.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusionestimation: addressing endogeneity in health econometric modeling. Health Economics , 27, 531-543.

Vella, F. (1993). A simple estimator for simultaneous models with censored endogenous regressors. International Economic Review , 34, 441-457.

Blundell, R.W., & Powell, J.L. (2004). Endogeneity in semiparametric binary response models. Review of Economic Studies , 71, 655-679.

Woodridge, J. M. (2002). Econometric analysis of cross section and panel data. Cambrige: MIT Press.

Wooldridge, J. M. (2009). Introductory Econometrics, A Modern Approach. South-Western Cengage Learning.

Zhang, X., Zhao, X., & Harris, A. (2009). Chronic diseases and labour force participation in Australia. Health Economics , 28, 91-108.

Zimmer, D. M. (2015). Employment effects of health shocks: The role of fringe benefits. Bulletin of Economic Research , 67, 346-358.

ABSTRACT

ESSAYS ON HEALTH AND LABOR MARKET OUTCOMES

by

MARYAM JAFARI BIDGOLI

August 2017

Advisor: Dr. Allen C. Goodman

Major: Economics

Degree: Doctor of Philosophy

This dissertation examines the relationship between health and labor market outcomes using the 1996-2010 longitudinal RAND Health and Retirement Study (HRS). First, it estimates the relationship between health and labor supply using a simultaneous equation model, treating health as endogenous. The effect of health may be overestimated because people may adjust their non-employment status by their health conditions (justification hypothesis). By using a full information maximum likelihood method, we can conduct a true test of exogeneity on the health variable, taking into account the correlation between two labor supply and health equations' time-varying error components (unobserved heterogeneity). The results confirm that health is endogenous to labor supply, and has a positive and significant effect on the males' (0.6833, $p<0.01$) and females' (0.6833, $p<0.01$) labor supply. The reverse effect of labor supply on health is also positive and significant (0.2981, $p<0.01$, males; 0.0305, $p<0.05$, females). The finding indicates that it is impossible to determine the direction of bias in the health effect for both males and females. Second, this dissertation examines the impact of health insurance coverage on the labor supply. To address the possible endogeneity of health insurance coverage to labor supply, I estimated the model for a group of married people who have spousal health insurance. The finding indicates that individuals with spouse's health insurance are more likely to exit the labor market (-0.5527, $p<0.01$, males; -0.7601, $p<0.01$,

females). Third, this study examines the short- and long-term impacts of cancer on the labor market outcomes for a sample of married people. The effect of cancer is negative and significant for those women cancer survivors who are diagnosed two years or fewer prior to interview (-0.08, $p<0.01$), and for those males who have survived for five years or more (-0.07, $p<0.10$). Employed men in the years immidiately following diagnosis, work 2.76 hours ($p<0.10$) less a week than other employed men. Employed women following three to five years since diagnosis, work 4.70 hours ($p<0.05$) less per week. Forth, using survival analysis, strong evidence was found that cancer survivors are more likely to experience a longer non-employment spell than a non-cancer group.

AUTOBIOGRAPHICAL STATEMENT

Maryam Jafari Bidgoli

Education

Ph.D. in Economics, Wayne State University, Detroit, MI. 2017
M.A. in Economics, Wayne State University, Detroit, MI. 2015
M.A. in Economics, University of Tehran, Tehran, Iran. 2008
B.A. in Economics, Islamic Azad University, Tehran, Iran. 2005

Fields of Concentration

Health Economics, Labor Economics, Health Insurance

Conference Presentations

- 2017 Michigan Academy Conference, Western Michigan University, Kalamazoo, MI. (March 10th, 2017).

Poster Presentations

- H2D2 Research Conference, University of Michigan, Ann Arbor, MI. (March 10th, 2017)
- 8th Annual Midwest Graduate Research Symposium, University of Toledo, Toledo, OH. (March 25th, 2017)
- WSU 2016 Graduate and Post-Doctoral Research Symposium, Wayne State University, Detroit, MI. (March 8th, 2017)
- WSU 2014 Graduate and Post-Doctoral Research Symposium, Wayne State University, Detroit, MI. (March 18th, 2014)

Awards and Honors

- Samuel M. Levin Award for Best Paper in Economics, Department of Economics, Wayne State University, Detroit, MI. (March 31st, 2017)
- Mendelson Award for Best Dissertation in Economics, Department of Economics, Wayne State University, Detroit, MI. (April 17th, 2017)
- Blue Cross Blue Shield of Michigan Foundation Student Award, 2016-2017 academic year
- Judging master's students posters at the Graduate and Postdoctoral Research Symposium, Wayne State University, Detroit, MI. (March 8th, 2017)
- Graduate Research Assistantship, Department of Psychiatry, School of Medicine, Wayne State University, Detroit, MI. (2014-2015)
- Graduate Teaching Assistantship, Wayne State University, Detroit, MI. (2010-2015)
- Persia House of Michigan (PHOM) 2014 Cultural (Merit) Award