The effects of breast cancer on individual labour market outcomes: an evaluation from an administrative panel*

Thomas Barnay[1,2], Mohamed Ali Ben Halima[3], Emmanuel Duguet[1,2,3], Christine Le Clainche[4,3] and Camille Regaert[5]

[1]Université Paris Est, ERUDITE (EA 437)
[2]FR TEPP CNRS 3435
[3]Centre d'Etudes de l'Emploi
[4]Université de Lille
[5]IRDES

January 2016

Abstract

Using an administrative data set (Hygie), we apply a difference-in differences with dynamic matching estimation method to the onset of breast cancer. The employment probability decreases by 10 percentage points (pp) one year after the onset of cancer compared to the not-treated group. The detrimental effect of breast cancer on employment significantly increases over time, up to 12 pp after five years. Our study also aims to identify some socio-demographic and work-related protective factors against adverse effects of breast cancer on labour market outcomes. We stress out four potential protective factors related to the negative breast cancer effect. Firstly, a young age at occurrence reduces this deleterious effect. Secondly, a high first job wage appears as a protective factor. Thirdly, having faced less unemployment in the past is associated with a weaker negative breast cancer effect on employment in the short run. Finally, we find a moderate "generation effect" after a stratification by year of cancer onset.

JEL: : I10, J21, J22.
Keywords: breast cancer, labour market participation, difference in differences, matching.

*Corresponding author: Thomas Barnay, Université Paris Est, Mail des Mèches, 61 avenue du Général de Gaulle, 94010 Créteil Cedex, France. thomas.barnay@u-pec.fr

1 Introduction

Thanks to advances in organized screening, detection and treatment, cancers may be related to chronic diseases (Cutler, 2008). The burden of breast cancer for the Statutory Health Insurance is important in France. This originates in the long term disease scheme which supports all the expenditures related to chronic diseases, including cancers. Therefore, the average cost of treatment for the breast cancers is 10 thousands euros per year in France. In 2012, the expenditures of all cancers reached 14 billion Euros including 2.3 billion for breast cancer alone. In addition, breast cancer causes indirect costs: lost days of work and productivity losses (Rapport de l'observatoire sociétal des cancers, 2014).[1] In the female population, the breast cancer onset occurs relatively earlier than other cancers [2] and needs treatments associated with functional sequels and therefore raises questions about the impact of breast cancer on individual well-being and especially on labour market outcomes. Furthermore, the net survival rate of women diagnosed with breast cancer between 1989 and 2004 in France, according to French cancer registries, is 97% after one year and 86% after five years (Jooste et al. 2013), allowing the study of the causes of work impairment.

Effect of cancer on labour market outcomes. According to health capital models (Grossman, 1972), the onset of cancer, like any serious health event, affects the career paths through the potential effects on the health stock, the decrease in productivity and in hours worked (Moran et al., 2011), the depreciation's rate of health capital and also the future investments in human capital. In a meta-analysis of 26 papers using US and European data, de Boer et al. (2009) estimate the relative risk of unemployment of cancer survivors to 1.37 in reference to a healthy population, all other things being equal. In a review of 64 international articles Mehnert (2011) shows that the average RTW (return to work) of cancer survivors increases from 40% six months after the diagnosis to 62% after twelve months, 73% after eighteen months and 89% after twenty-four months. Conversely, the onset of cancer also motivates a permanent exit from the labour market. In Finland, after accounting for age and gender differences, the relative risk of early retirement is 2.2 for survivors of cancer of the nervous system, 2 for leukemia, 1.9 for tongue, 1.2 for breast and 1.1 for prostate (Taskila-Abrandt et al., 2005). In a previous French study, we show that the onset of cancer decreases the probability of being employed by 7-8 percentage points (one year after the cancer onset) and 13 (five years after), in the same proportions for male and female populations (Barnay et al., 2015). The negative impact of cancer on the career path passes mostly through functional limitations (Bradley et al., 2002, USA) which may be specific, such as arm pain for breast cancer as a major sequel of treatment (Quinlan et al., 2009 in Canada, Blinder et al., 2012, in the USA) as well as depressive episodes (Damkjaer et al., 2011, in Denmark) and memory and concentration disorders (Oberst K. et al., 2010, USA). These effects are amplified or attenuated depending on the nature of the initial endowments of human capital, the difficulty of pre diagnosis working conditions but also the type of cancer (site, severity of the disease) and finally, the nature of the treatment (Mujahid et al., 2011; Lindbohm et al, 2011; Johnsson et al., 2011; Blinder et al., 2012). Past professional biography (unemployment or training episodes) can also lead to stigmatizing effects on the careers of individuals (Heckman and Borjas, 1980; Gregg and Tominey, 2005) and, for some social groups,

[1] 153 000 new cases of breast cancer were estimated for 2011. The death rate from breast cancer is 22.3 per 100 000 inhabitants in France, close to the OECD average is 20.1.

[2] The median age at breast cancer onset is about 65 years in France in 2012 (INCA 2015). Cancer before 40 represents 5% of the cases diagnosed. Diagnosis of new cases are generally made after 55 years old, partly due to systematic screening from 50 years old (Inca, 2014).

predicts the occurrence of professional cancers. Feuerstein et al. (2010) thus stress the importance of improvements in the workplace in terms of schedule flexibility, social support from colleagues, social climate and job stress in order to protect work in cancer survivors.

Focus on Breast cancer. A large international literature (in particular with US data) is devoted to effect of breast cancer on professional path (Bradley et al., 2006 ; Bradley et al, 2013 ; Chirikos et al., 2002a; Chirikos et al., 2002b; Drolet et al., 2005; Heinesen and Kolodziejczyk, 2013). For instance, Bradley et al. (2007) show that the negative effect of cancer on employment lasts significantly 6 months after the diagnosis but not more. Heinesen and Kolodziejczyk (2013) measure causal effects of breast and colorectal cancer on labour market outcomes. On the basis of Danish administrative data, they estimate the ATT (average treatment effects on the Treated) by propensity score methods using persons with no cancer as a control group. Suffering from a breast cancer in year t reduces the probability of being employed by 4.4 percentage points in year $t+1$, by 5.7 percentage points in year $t+2$, and 6.7 percentage points in $t+3$. From Danish data (2001-2009), Carlsen et al. (2014) stress that women, after a breast cancer, who experienced periods of unemployment before the diagnosis have an increased risk of being unemployed after in reference to women at work before diagnosis (79 weeks of unemployment against 26 weeks for working women before diagnosis). Past French studies are more limited. Eichenbaum-Voline et al. (2008) and Joutard et al. (2012) carried out a matching method on survey data which includes treatment variables. Marino et al. (op. cit.) show that two years after the diagnosis of cancer the probability of returning to work in the female population is 72% (against 25 % six months after the diagnosis).

Many articles underline the role of different health and socioeconomic characteristics which influence the effect of breast cancer on employment. At first, a significant body of the litterature points out the nature of cancer and types of treatment (Jagsi et al., 2014; Hassett et al., 2009). Treatments require an exit from the labour market which may be long when women undergo a combination of treatments (surgery, radiotherapy, and chemotherapy). In France as in other developed countries, the most detrimental effect regarding the delay according to which a woman with breast cancer can return to work, is a combination of treatments and especially chemotherapy before or/and after radiotherapy (for France with a data set that includes the severity of the disease and the type of treatments, see Duguet, Le Clainche 2016). Women who have undergone a surgery with partial mastectomy followed by radiotherapy can often return to work in the 6 months following the surgery, if no comorbidity occurs.

Treatment and comorbidities factors, sociodemographic and work-related characteristics may also explain differences in terms of Labour outcomes (Bradley et al., 2004, Torp et al., 2012). The differences between countries regarding the delays beyond 6 months may be explained by protective factors such as favourable social protection rules (sick leave legal rules, social insurance schemes, work flexibility) and the way the working conditions can be implemented by the employers. Using French data, Duguet and Le Clainche (2016) show that the probability of returning to work two years after the diagnosis, and especially for women after a breast cancer, increases when appropriate working conditions are implemented. By and large, the onset of cancer affects future investments in human capital (primary or secondary health prevention) due to the difficulty to combine work and cancer treatment (Yarker et al., 2010; Johnsson et al., 2010). The onset of cancer can also modify the nature of the labour contract (e.g. full-time/part-time, working hours). Many studies shed light on the relationship between cancer occurrence and work duration (Petersson et al., 2011; Farley et al., 2008; Torp et al., 2012; Paraponaris et al., 2010). From a Swedish sample of 756 working women who have un-

dergone a breast cancer surgery, Petersson et al. (2011) find that, one month after the surgery, 56% of women with breast cancer are on sick leave, the majority for full-time. Most of them are employed at diagnosis and 91% of those work greater than 75% of full-time. According to the study of Farley et al. (2008), in the USA for both genders survivorship affected the probability of working full-time and hours after 2-6 years post-diagnosis. In the female population of survivors after a new cancer, the cancer effects are 14 to 17 pp for the employment rate, 14 to 18 pp for full-time and 7 to 8 for hours per week. Torp et al. (2012) from a Norwegian database highlight that a low socioeconomic position appears as a risk factor for returning to work. On the basis of Korean data (1993-2002), working women after a breast cancer diagnosis are more often unemployed if they have a low education or a low income (Eunmi et al., 2009).

From French data, Paraponaris et al. (2010) study the relationship between the cancer occurrence and type of labour contract. Their findings indicate that fixed-term contracts are at greater risk of job loss for workers in the female population (-8 pp in reference to permanent contract).

From a theoretically point of view, the return to work depends on economic incentives. Bradley et al. (2013) show that the negative effect of breast cancer on employment is reduced if the health insurance depend on the job. This result refers to the "job lock" assumption e.g. workers remain in their current job in order to maintain their health insurance. In contrast to the USA, in France, for particular diseases which need intensive, expensive and long-term care (such as cancer), the long-term disease scheme has been implemented in order to pay for the related treatment costs and to provide an equal access to health care.

In this study, we focus on the sociodemographic and work-related sides of cancer. Due to the French legal frame, we cannot assume a job lock effect. We can expect that the negative effect of breast cancer will last over time at least if the woman does not suffer from a recurrence of the disease, as in Heinesen and Kolodziejczyk (2013) in the Danish case. We examine four potential types of socioeconomic factors that could influence employment in a short-long. At first, a young age of the breast cancer's occurrence could be less disadvantageous than at a later age (Petersson, 2011). Secondly, a good socioeconomic position is associated with appropriate working conditions following diagnosis and also with a better return to work (Eunmi A. et al., 2009). Third, we account for the stability of the past career before the onset of cancer (Heckman and Borjas, op.cit.; Gregg and Tominey, op.cit.). Finally, we also examine a generation effect. The medical progress coud lead to a better return to work for the last generation of cancer survivors but other factors, such as discrimination[3] on the Labour Market, could also play a more ambiguous role.

Using panel data from the National Pension Fund and the National Health Insurance Fund we examine two issues. First, we estimate, for the first time in France, the effects of breast cancer up to five years after its onset on employment outcomes. We perform a difference-in-differences analysis combined with a dynamic matching algorithm. Second, we highlight the role of protective factors which attenuate the adverse effects of cancer on labour market outcomes. The paper is organised as follows. Section 2 is devoted to the methodology and the econometric model specifications. In section 3, we present our main findings and a discussion is provided in section 4. The last section presents our conclusions.

[3] Paraponaris et al (2010) pinpoint the endogeneity of discrimination for women survivors after a cancer. They underline that productivity or number of children could contribute to decrease the RTW through the effect of the feeling of discrimination.

2 Data and methodology

Data. The HYGIE data set has been constructed from two nationwide administrative sources. The HYGIE data were extracted from the National Pension Fund (CNAV) and the National Health Insurance Fund for Salaried Workers (CNAMTS) administrative databases. The resulting database contains individual information regarding beneficiaries, their professional careers, medical consumption, sick leaves, employees' professional context and a few characteristic concerning employer establishments. The HYGIE database makes it possible to study the relationships between health, work, professional career and firm characteristics. HYGIE is representative of private sector employees in France.

More specifically, the files were extracted from the National Career Management System (SNGC), which groups all of the private sector employees in France, and the National Statistical Beneficiaries System (SNSP), which groups all of the private sector retirees in France, matched with sickness benefit data taken from the National Health Insurance Inter-regime Information System (SNIIR-AM). The CNAV data constituted the point of entry and included a random sample of beneficiaries aged 22 to 70 years old who contributed to the general pension fund at least once in their lives. The CNAMTS data concerns both primary and secondary beneficiaries of the National Health Insurance scheme who received sickness benefits for at least one health service in 2003, 2004 or 20052. The linkage of the CNAV and CNAMTS data enabled us to build the HYGIE database panel of 538,870 beneficiaries from 2005 to 2010.

There are 552,048 working people in this data source. The restriction to women leaves 225,340 observations. For this study, we use a data set that includes the whole career of the workers from their first job to year 2008. We drop the retirement years and keep the activity period only in order to evaluate the impact of cancer on employment. The demographic data include gender and the birth year. The HYGIE data include the wage of the first job. We take the ratio of this first job wage to the yearly median wage and compute four equal sized starting wage classes. This way to proceed corrects for inflation and gives the position of the worker on the income ladder.

The medical data include a sick leave dummy and the International Statistical Classification of Diseases and Related Health Problems (henceforth ICD). We identify breast cancer with the ICD code C50. In order to perform the dynamic matching methodology, we keep all the women in the sample. That is: women with a breast cancer, women with no disease at all and women with another disease. These other diseases include all the cancer types but also the French definition of long term diseases (like diabetes).

These diseases are on an official list, and must last at least 6 months. They are fully financed by the statutory health insurance. This way to proceed also guarantees that the estimation results obtained for all the cancers and diseases in this data set are comparable, since the same data set can be used for all of them.

Three labour market statuses were identified in the HYGIE data: employment, regular employment and unemployment. The Outcome variables giving the individual's employment status were identified in the HYGIE database on the basis of contributions paid for qualifying periods (unemployment, sickness/occupational accident) and quarters of national social security scheme contributions. Individuals were thus identified as follows. Regular employment means at least one quarter of contribution to the national social security and the absence of unemployment spells. Employment means at least one quarter of contribution, with or without unemployment spells.

We use the following indicator of career stability. We divide the number of years with an un-

employment spell by the total number of years spent in the labour market. We match women on the value of this indicator one year before the cancer onset, so as to control for their past performance in the labour market. We use the same type of indicator for past health: the number of years with sickness leaves divided by the total number of years spent in the labour market. We also use the lagged value of this indicator (one year before cancer) when we match women.

In order to perform the dynamic matching methodology, we keep all the women in the sample. That is: women with a breast cancer, women with no disease at all and women with another disease. These other diseases include all the cancer types and also all the diseases that correspond to the French definition of long term diseases (like diabetes). These diseases are on an official list, and must last at least 6 months. They are fully financed by the compulsory health insurance. This way to proceed also guarantees that the results obtained for all the cancers and diseases in the data set are comparable, since the same data set can be used for all of them.

Methodology. Our estimator is similar to Heckman et al. (1997). We wish to explore the impact of breast cancers on the employment history of women. We account for four problems. Firstly, it is likely that cancer can cause a break in individual employment histories. In this case, only a dynamic approach can identify the break with a before-after analysis, where the break date is the date of the cancer which is specific to each individual. A static approach can only compare individuals with a different health status at the date of the survey, and cannot analyse directly the impact of cancer for each individual. Secondly, the women in the data set have different ages so that the observation window differs from one woman to another. Therefore we need to match women by age, so that we compare women that had a health event (or not) at the same age and during the same year. Thirdly, we account for the panel data model correlated individual and time effects. We also allow for non parallel time effects by adding matching to the usual difference in differences estimator. Fourthly, we perform a non parametric estimation, so that no specific distributional assumption is made.

In the standard difference in difference approach (Henceforth DiD, see Lee, 2003 for an exposition), the variables that are constant over time should not alter the estimations so that we could have used this approach that controls for individual unobserved heterogeneity. Our approach generalizes DiD in two directions. First, we match on the lagged endogenous variables so that we compare women that had reached the same situation in the labour market before the health event. This allows for a better assessment of the health-labour causality since the women in our cancer group have the same occupation than the women in the control group before the health event occurs. Therefore, we can rule out the reverse causality from labour to health, since the women that work more often are matched with similar women in the control group. Second, the standard DiD method assumes that the time effects share parallel evolutions in the health-event group and in the control group. By matching on individual characteristics we allow for the time trends to be different in the health and control groups, as far as the slopes of the trends depend on the individual characteristics and on the lagged endogenous data. For these reasons, our application of DiD should provide a more robust estimation than in the standard case.

The Hygie data set provides a detailed dynamic account of two main variables: cancer occurrence and occupational status. In this section we analyse how it is possible to evaluate the impact of cancer on the subsequent labour market history. In order to identify successfully the impact of health events we need to account for two types of quantities: on the one hand, the difference in histories between the women that experienced cancer and the other women; on the other hand, the history variation of one woman before and after cancer. In this section, we

show that the difference-in-differences method with matching allows us to estimate the effect of cancer by controlling both for the observable individual variables and the non-observable individual heterogeneity, including when it is correlated to the observable individual variables.

The outcome variables are the annual activity dummies corresponding to the three following occupations: employment, unemployment and sick leave. One can interpret our analysis as an assessment of the impact of the breast cancer on these occupational dummies.

We consider all the women with a breast cancer, denoted $i \in I$, where I stands both for the index set of the women with a breast cancer and their number. A woman $i \in I$ is observed between the years t_i^- and t_i^+ and a breast cancer happens on year $t_i \in (t_i^-, t_i^+)$. In order to evaluate the effect of the cancer, we compare the occupation of woman i in $t_i - 1$ to the occupational choice k years after the health event, in $t_i + k \leq t_i^+$. In what follows, we take employment as example but any other occupation can be dealt with in the same way. The employment probability of woman i during year t, denoted $p_{i,t}$, depends on a vector of individual explanatory variables X_i, an unobservable individual effect α_i, potentially correlated with X_i, a time effect $\beta_{0,t}$ and a joint effect of the explanatory variables with the time effect $\beta_{1,t}(X_i)$. The employment dummy variable $d_{i,t}$ follows a Bernoulli process with mean $p_{i,t}$ given by:

$$\begin{aligned} d_{i,t} &= p_{i,t} + \epsilon_{i,t} \\ p_{i,t} &= f_i(X_i) + \alpha_i + \beta_{0t} + \beta_{1,t}(X_i) + \gamma_i(t - t_i) \times T_{i,t} \end{aligned}$$

where $f_i(.)$ is an unknown function relating X_i to the employment probability $p_{i,t}$, γ_i is the effect of the breast cancer on the probability to be employed and $T_{i,t}$ a dummy variable equal to 1 if there is a breast cancer ($t \geq t_i$), 0 otherwise ($t < t_i$). The γ_i terms depend on how much time passed since the breast cancer occurred $t - t_i$. The $\epsilon_{i,t}$'s are idiosyncratic error terms with $\mathrm{E}\left(\epsilon_{i,t} \middle| X_i, \alpha_i, \beta_{0t}, \beta_{1t}(X_i), T_{i,t}\right) = 0$. Henceforth, we consider the effect of the cancer between $t_i - 1$ and $t_i + k$, so that we wish to estimate an average value for $\gamma_i(k),\ k \geq 1$.

The estimation proceeds through the elimination of all the components but $\gamma_i(k)$. The techniques used to achieve this goal are based on differencing (for α_i and $\beta_{0,t}$), matching (for X_i and $\beta_{1,t}(X_i)$) and averaging (for $\epsilon_{i,t}$). In the first step, we will match the women experiencing a cancer ($i \in I$) with their twins defined as:

$$J(i) = \left\{ j : t_j^- \leq t_i - 1, t_i + k \leq t_j^+,\ t_j > t_i + k \text{ and } X_j = X_i \right\}$$

the two first inequalities simply impose that the twins should be present over the same period than woman i. The third inequality defines dynamic matching, the twins $J(i)$ should experience their cancer (or another long term disease, if they have any) after the end of the comparison period of woman i. This implies that we match i with, on the one hand, women that will neither experience cancer nor another long term disease and, on the other hand, women that will experience either cancer or another long term disease at a later date. When somebody does not experience cancer (or another long term disease), we use the convention $t_j = \{+\infty\}$. Eventually, the twins should have the same individual characteristics. The notation $J(i)$ will also be used to indicate the number of twins of woman i. Notice that two women can share common twins, since we make use of all of them for each woman. The outcome variable of the twins does not include the effect of the cancer by definition, so that their outcome variable is given by:

$$\begin{aligned} d_{j,t} &= p_{j,t} + \epsilon_{j,t} \\ p_{j,t} &= f_j(X_j) + \alpha_j + \beta_{0t} + \beta_{1,t}(X_j) \end{aligned}$$

and the average outcome of i's twins is given by:

$$\frac{1}{J(i)}\sum_{j\in J(i)} d_{j,t} = \beta_{0t} + \beta_{1,t}(X_i) + \frac{1}{J(i)}\sum_{j\in J(i)}\left(f_j(X_i) + \alpha_j + \epsilon_{j,t}\right)$$

Consider first the difference between woman i and all the twins $j \in J(i)$ before the health event, we eliminate the terms in β_{0,t_i-1} and $\beta_{1,t_i-1}(X_i)$ and get:

$$\begin{aligned} D_{i,t_i-1} &= d_{i,t_i-1} - \frac{1}{J(i)}\sum_{j\in J(i)} d_{j,t_i-1} \\ &= f_i(X_i) + \alpha_i + \epsilon_{i,t_i-1} - \frac{1}{J(i)}\sum_{j\in J(i)}\left(f_j(X_i) + \alpha_j + \epsilon_{j,t_i-1}\right) \end{aligned} \tag{1}$$

and when we take the difference after the cancer date we also eliminate the β components:

$$\begin{aligned} D_{i,t_i+k} = &= d_{i,t_i+k} - \frac{1}{J(i)}\sum_{j\in J(i)} d_{j,t_i+k} \\ &= f_i(X_i) + \alpha_i + \gamma_i(k) + \epsilon_{i,t_i+k} - \frac{1}{J(i)}\sum_{j\in J(i)}\left(f_j(X_i) + \alpha_j + \epsilon_{j,t_i+k}\right) \end{aligned} \tag{2}$$

the difference in the differences (2) and (1) therefore leads to:

$$\begin{aligned} DD_i(k) &= D_{i,t_i+k} - D_{i,t_i-1} \\ &= \gamma_i(k) + \epsilon_{i,t_i+k} - \epsilon_{i,t_i-1} - \frac{1}{J(i)}\sum_{j\in J(i)}\left(\epsilon_{j,t_i+k} - \epsilon_{j,t_i-1}\right) \end{aligned}$$

so that $\mathrm{E}(DD_i(k)) = \gamma_i(k)\ \forall i,k$. Our estimator is simply the average of these individual health effects. We define:

$$\hat{\gamma}(k) = \frac{1}{I}\sum_{i\in I} DD_i(k)$$

so that:

$$\mathrm{E}\left(\hat{\gamma}(k)\right) = \frac{1}{I}\sum_{i\in I}\gamma_i(k).$$

It remains to compute the variance of $\hat{\gamma}(k)$. The method is explained in appendix. We performed the estimation with SAS 9.4.

3 Results

Our sample includes 2,547 women who suffer from a breast cancer with enough years to compute a before-after difference. These women will be matched with 203,392 women with no long term (six months) disease and 19,059 women with another long-term disease. Overall, women which experienced breast cancer have similar relative revenues than the others. They are a little less present in the highest revenue class. They are also significantly older, as expected. In 2008, 60.4% of women with no disease were less than 45 years old, while only 13.5% of women with a breast cancer were (Table 1).

We use a matching method in order to eliminate the effect of the confounding variables on this naive average outcome difference. We then perform five estimations (full sample; by age of

cancer onset (more or less than the median age 48); by first wage classes (in quartiles), by stability of professional career (measuring from the indicator (c) which represents the share of the number of years worked with at least one unemployment spell on the number of years worked ; three classes are selected $c = 0$, low unemployment ($0 < c \leq 0.16$) and high unemployment ($c > 0.16$) and by onset year with three classes ($t \leq 1990$, $1991 \leq t \leq 1999$ and $t \geq 2000$) allowing to measure short (one year) and long term (from 2 to 5) breast cancer effects on two labour market outcomes (employment, regular employment e.g. without unemployment spells).

Effects on employment. We use two indicators of employment. The first indicator is a dummy variable equal to one when a worker has validated at least one quarter in employment. We call it "employment" because it is representative of a job drawn at random in the total population. This definition includes all types of job: stable jobs, but also insecure jobs and some quarters of unemployment or disease in the same year. In order to better assess the effect of breast cancer, we also use a more restrictive definition of employment "regular employment". Here we impose that there is no unemployment during the same year. Notice that both definitions of employment are compatible with fixed term contracts, so that the "regular employment" jobs need not be especially stable over time. This indicator just signals workers that have been working during the year without knowing unemployment.

The employment probability decreases by 10 percentage points (pp) one year after the onset of cancer compared to the not-treated group. We also observe that the detrimental effect of breast cancer on employment significantly increases over time to reach 12.4 pp five years later. If we consider regular employment (without unemployment spells), the adverse effect remains but is less pronounced (from 6.3 pp in $t+1$ to 10.5 in $t+5$). At this stage, we can assume that the regular employment variable selects a protected population since it copes better with the negative effects of a breast cancer on the career path.

A young age of occurrence. We compare the effect of cancer depending on the age of the worker at the moment of breast cancer. For workers under 48 (the median age at cancer in our data), the negative impact of a breast cancer remains constant over time (between 8 and 10 pp for irregular employment; at 7 pp for permanent employment), whereas it linearly grows in the older female population (above 48) respectively from 8 pp to 15 pp for employment, and from 5 to 13 pp for regular employment. These patterns suggest that a late cancer onset is more detrimental to the employment of breast cancer survivors, with a negative effect that increases over time. This pattern should be explored in depth and is probably due to opposite effects. Indeed, from a medical point of view, breast cancers at a younger age (before menopause) are known to be more aggressive, everything being equal, than a cancer onset occurring after the menopause (most of time after the mean age of 51 in France).[4] All at once, early cancers often diagnosed with delays due to the absence of screening programs and low prevalence (less than 5% of breast cancers occur before 40). Then early cancers usually receive a worse prognosis. Furthermore, we can assume that the young women may have more "energy" to return to work earlier than older women, which tend also to be more from discriminated against. Otherwise several other explanations might account for this age at cancer consequences. First, the nature of co morbidities and treatments might differ depending on the

[4] Cancers at a young age are more often "triple negative" which prevents the use of efficient treatments. Otherwise, hormonotherapy for hormonodependent cancers treatments are more efficient after menopause than before and the neo adjuvant therapy for non hormonodepent cancers is not considered as really efficient (see eg Pourquier, 2000).

age of occurrence. Second, breast cancer occurring at older ages can be especially disabling and prevent women from keeping a job (undergoing an hormonotherapy which is more often targeted on post menopaused women is given for five years and is often provoking unpleasant side effects). Third, as more women approach the retirement age, the opportunity cost of exiting employment decreases. Fourth, the decline in the probability of employment for older women may be explained by a phenomenon of "double penalty" that can lead to amplified effects of exit from the labour market. Traditional analyses of investments in human and health capitals can be enhanced by the phenomena of changes in preferences or discrimination related to age that has been raised especially for older workers (e.g. Datta Gupta and Larsen, 2010).

High income effect. We test the protective role of the starting income on the negative impact of breast cancer on employment by dividing the sample into four classes (relative wage at the date of entry in the labour market).[5] Our findings show that the upper relative wage class workers (above the third quartile of the relative wage) are especially protected. The negative effect of cancer varies from 3 to 12 pp only, while it is much bigger for the other workers (for both employment measures). The decrease of the employment rate in the lowest relative wage class tends to increase over time from 10 pp ($t+1$) to 18 pp ($t+5$). This detrimental effect of cancer widens over time for all wage classes. The effect of cancer is strictly decreasing with the relative wage class from the first to the third year after the onset of cancer. For instance for employment, the effects after one year are -15 pp, -12 pp, -9.3 pp and -4.7 pp). Several arguments can explain these results. First, The French social protection system is highly redistributive and allows lower income workers to get a better income replacement rate than the higher income workers. This may create an inactivity trap that would explain a long term effect of -17 pp for the first quartile). Second, the lower wage class workers are more often concerned with co-morbidities associated with poor working conditions, and tend to exit the labour market permanently. Third, the lowest quartile workers cannot benefit easily from work arrangements because the nature of their jobs makes it difficult.

A career stability. We test another assumption related to the nature of the past career (measured by the occurrence of unemployment spells in past years). The results show that employment and regular employment both decrease by 5 pp one year after a breast cancer for women without any past unemployment spell. This short-term effect is higher for those who were unemployment in the past. Workers with a low past unemployment ratio face a 11 pp decrease in employment, while people with a strong past unemployment ratio (above 16%) face a 16.6 pp decrease. However, this result does not extend to regular employment, since all the workers with past unemployment spells face a 8 pp decrease one year after the onset of cancer. If we consider a longer time horizon, comparison of effects over a longer time horizon shows contrasted findings according to the stability class. People with a low past unemployment rate face a stronger decrease in their regular employment rate that the workers in the highest past unemployment rate. This could indicate the following dynamics: people with a low past unemployment rate would be driven toward unemployment more often and move in a higher unemployment class in the future. Our findings pinpoint that for individuals who had previous unemployment spells, it is more difficult to remain in regular employment after a cancer shock.

[5] The relative wage is equal to the ratio of the entry wage to the median entry wage of the same year. Four classes are defined by the quartiles of this variable that is defined for all workers. Notice that the quartiles are not computed on the cancer population but on the total population.

A generation effect. The last test deals with a potential generation effect. Our idea concerns the expected positive medical progress on the return to work for the young generation. In order to test this hypothesis, we perform the estimation by onset year and consider short and medium terms (from $t+1$ to $t+3$). The year following the diagnosis, the fall of irregular employees is very similar whatever the onset year (about 9-10 pp) that seems to overturn the innovation assumption. Nevertheless, we stress out a clear generation effect concerning effect on regular employment. For a cancer occurred after 2000, the decrease in employment represents 5.2 pp versus 6.2 pp between 1990 and 2000 and 9.3 pp if the cancer appears before 1990. The gap between generations is corroborated for two and three years after the cancer onset. The decline in employment for the new generation can be explained by several factors. The improvement in cancer medical treatments during recent years can support this effect. The incidence of in situ cancers increased significantly from 1990 to 2005 in all age groups, but particularly among 50 to 74 years women. This trend reversed after 2005. The incidence of invasive cancers grew slowly from 1990 to 1996 and then more sharply in 1996, mainly among women 50-74 years before declining in 2004. Finally, cancers advanced stage at diagnosis decreased after a peak incidence reached in the early 2000s. Incidence developments probably reflect the combined action of several factors (risk factors, screening and diagnostic techniques). In addition, organized screening mammography breast cancer was widespread in France in 2004. This program allows all 50 and 74 years old women to have a mammogram and a clinical breast exam once every two years. Thus, for breast cancer, the 5-year survival increased from 81% in 1990 to 89% in 2002.

4 Discussion and conclusion

The originality of our study involves several fields. First, the estimation of the cancer effect on professional situations covers a long-term period from one to five years after the registration of Cancer. In addition, it relies on administrative data for identifying the careers of a large sample of private sector employees. Then, the sample size permits performing a DiD with exact matching and to define a rigorous control group that exploits the panel dimension of the data. Finally, we examine a relatively young female population (in our sample, 43% of the women with a breast cancer are less than 45), which reinforces the relevance of the analysis of professional paths of this population.

Our main findings confirm, for the first time in France with this econometric method, the detrimental effect of breast cancer on employment. The proportion of individuals who have completed at least one quarter of employment strongly decrease after the onset of cancer. By 10 pp after one year, and the effect is long-lasting since it is still 12 pp five year after the cancer onset.

It is obviously difficult to comparer our results with the ones from the other studies because of the differences in the data, methods, and also international differences in labour market structure, public financing of cancer's cost and sick leave. Keeping this in mind, our findings are close to Heinesen and Kolodziejczyk (2013) and Moran et al. (2011), who use similar econometric methods by combining DiD and the propensity score. From Danish administrative register data, Heinesen and Kolodziejczyk (2013) estimate effects of breast cancer on labour market outcomes for threeyear survivors. With both different group controls, they find following probabilities of employment's decrease; 4.4 pp in $t+1$, [5.3-5.8] in $t+2$ and [6.2-6.7] in $t+3$. On US data, Moran et al. (2011) focus on a young population, as our study does (cancer at younger ages: 28-54 years) and estimate the effect of surviving cancer on long-term employ-

ment outcomes (2-6 years post-diagnosis). Breast cancer survivors had employment rates that were 7-8 percentage points lower than the group control. These findings are significantly lower than ours. In general, the generosity of the French health insurance system can explain a more frequent trade-off in favour of inactivity.

The originality of our study is to focus on many demographic and professional characteristics as protective factors against the deleterious effects of breast cancer on employment. As expected, a young age of occurrence, a high starting wage class and past employment stability promote a better return to work (close to Heinesen and Kolodziejczyk results in 2013). Many studies have pointed out social gradients in breast cancer survivors (e. g. Carlsen et al., 2008). Nevertheless, the most interesting findings deal with the potential generation effect related to the medical progress. This first result should be deepened.

Finally, several limitations should be reported. Besides that related to the definition of cancer (specific to this study), the data do not allow to identify the cancer stage, the type of treatment or the severity. Before the Cancer Plan which has been adopted for the first time in France in 2004, that is during the common period of our study, an hypothesis can be made regarding the existence of a link between aggressiveness of treatments (that is with hard pain and potential bad long term side effects) and social gradients. Indeed, due to better prevention behaviours, upper revenue class women are often better diagnosed and at an early stage of their cancer and so can recover without too much sequels. However, this holds only for the older women (who are also benefiting from organized screening cancers programs).

References

[1] Barnay T, Ben Halima M.A, Duguet E, Lanfranchi J et Le Clainche C (2015), "La survenue du cancer : effets de court et moyen termes sur les situations professionnelles", Economie et Statistique, No 475-476, pp. 157-186.

[2] Barnay T. (2015) "Health, Work and Working Conditions: A Review of the European Economic Literature", European Journal of Health Economics, DOI: 10.1007/s10198-015-0715-8. First Online.

[3] Blinder V. S., Patil S., Thind A., Diamant A., Hudis C. A., Basch E. et Maly R. C. (2012), "Return to work in low-income Latina and non-Latina white breast cancer survivors: a 3-year longitudinal study", Cancer, vol. 118(6), pp. 1664-1674.

[4] 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", Journal of Health Economics, 32(5), pp. 833-49.

[5] Bradley C.J., Neumark D, Luo Z, Bednarek HL. (2007), "Employment-contingent health insurance, illness, and labor supply of women: evidence from married women with breast cancer", Health Economics, 16(7), pp.719-37.

[6] Bradley C.J., Oberst K., Schenk M. (2006), "Absenteeism from work: the experience of employed breast and prostate cancer patients in the months following diagnosis", Psychooncology, 15 (8), pp. 739-747.

[7] Bradley C. Neumark D, Bednarek HL, Schenk M. (2004) : "Short-term effects of breast cancer on labor market attachment : results from a longitudinal study", Journal of Health Economics, 24, pp.137-160

[8] Carlsen K., Badsberg JH., Dalton S. (2014), "Unemployment among breast cancer survivors", Scandinavian Journal of Public Health, 1-10.

[9] Chirikos T. N., Russell-Jacobs A. et Jacobsen P. B. (2002a), "Functional impairment and the economic consequences of female breast cancer", Women and Health, 36 (1), pp.1-20.

[10] Chirikos T. N., Russell-Jacobs A., Cantor A. B. (2002b), "Indirect economic effects of long term breast cancer survival". Cancer Practice, 10 (5), pp. 248-255.

[11] Cutler, D.M., (2008), "Are we finally winning the war on cancer?", Journal of Economic Perspectives, 22 (4), 3-26.

[12] Damkjaer, L. H., Deltour, Suppli I., Palm N., Kroman J., Johansen N.T., Dalton C., Oksbjerg S., (2011), "Breast cancer and early retirement: Associations with disease characteristics, treatment, comorbidity, social position and participation in a six-day rehabilitation course in a register-based study in Denmark", Acta oncologica, 50(2), pp. 274-81.

[13] De Boer AG, Taskila T, Ojajarvi A, van Dijk FJ, Verbeek JH., (2009), "Cancer survivors and unemployment : a meta-analysis and meta regression", JAMA, 301(7), pp.753-62.

[14] Drolet M., Maunsell E., Mondor M., Brisson C., Brisson J., Masse B., Deschenes L. (2005), "Work absence after breast cancer diagnosis: a population-based study", Canadian Medical Association Journal, 173 (7), pp. 765-771.

[15] Duguet E., Le Clainche (2016), "Une évaluation de l'impact de l'aménagement des conditions de travail sur la reprise du travail après un cancer", Revue Economique, 71(1), pp. 49-79.

[16] Eichenbaum-Voline S., Malavolti L., Paraponaris A., Ventelou B. (2008), "Cancer et activité professionnelle", La Revue de l'OFCE, 104, pp. 105-134.

[17] Eunmi A. et al. (2009) : "Impact of breast cancer diagnosis and treatment on work-related life and factors affecting them", Breast Cancer Research and Treatment, 116 (3), pp. 609-616.

[18] Farley P., Vasey J. and Moran J. (2008), "Long-Term Effects of Cancer Survivorship on the Employment of Older Workers", Health Services Research, 43(1), pp.193-210.

[19] Feuerstein M., Todd B. L., Moskowitz M. C., Bruns G. L., Stoler M. R., Nassif T. et Yu X. (2010), "Work in cancer survivors: A model for practice and research", Journal of Cancer survivorship, research and practice, 4, pp. 415-437.

[20] Gregg P., Tominey E. (2005), "The Wage Scar from Youth Unemployment", The Centre for Market and Public Organisation 04/097, Department of Economics, University of Bristol, UK.

[21] Grossman M. (1972), "On the Concept of Health Capital and the Demand for Health", Journal of Political Economy, 80 (2), pp. 223-255.

[22] Hassett M., O'Malley J., Keating N., (2009), "Factors influencing changes in employment among women with newly diagnosed breast cancer", American Cancer Society, 115, pp.2775–2782.

[23] Heckman J.-J., Borjas G. (1980), “Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence”, Economica, 47, pp. 247-283.

[24] Heckman, J.J., Ichimura, H., Todd, P.E. (1997), “Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme”, Review of Economic Studies, 64 (4), pp. 605–654.

[25] Heisenen E., Kolodziejczyk C. (2013), “Effects of a Breast and Colorectal Cancer on Labour Market Outcomes-Average Effects and Educational Gradients”, Journal of Health Economics, 32, pp. 1028-1042.

[26] Jagsi, R, Hawley S., Abrahamse P., Li Y., Janz N., Griggs J., Bradley C., Graff J., Hamilton A., Kratz S. (2014), “Impact of adjuvant chemotherapy on long-term employment of survivors of early-stage breast cancer”. Cancer 120, (12), pp. 1854-1862.

[27] Johnsson A., Fornander T., Rutqvist L., E., OlssonM. (2011), “Work Status and Life Changes in the First Year after Breast Cancer Diagnosis”, Work, 38, pp. 337-346.

[28] Joutard X., Paraponaris A., Sagaon-Teyssier L., Ventelou B (2012). “A Continuous-time Markov Model for Transitions Between Employment and Non-employment: the Impact of a Cancer Diagnosi”", Annals of Economics and Statistics, vol. 107-108, pp.239-266.

[29] Jooste V., Grosclaude P., Remontet L., Launoy G., Baldi I., Molinie F., Arveux P., Bossard N., Bouvier AM., Colonna M. (2013), “Unbiased extimated of long-terme net survival of solid cancers in France”, International Journal of Cancer, 132(10), pp. 2370-7.

[30] Lindbohm M.-L., Kuosma E., Taskila T., Hietanen P., Carlsen K., Gudbergsson S. et Gunnarsdottir H., (2011), “Cancer as the cause of changes in work situation”, Psychooncology, 20, pp. 805-812.

[31] Marino P., Sagaon Teyssier L. , Malavolti L., Le Corroller, Soriano A. G. (2013), “Sex Differences in the Return To Work Process of Cancer Survivors 2 Years After Diagnosis : Results From A Large French Population-Based Sample”, Journal of Clinical Oncology, 31, pp.1-10.

[32] Mehnert, A. (2011), “Employment and work-related issues in cancer survivors”, Critical reviews in oncology Hematology, 77, pp.109-130.

[33] Moran J.R., Short P.F., Hollenbeak C.S. (2011), “Long-term Employment effects of Surviving Cancer”, Journal of Health Economics, 30 (3), pp. 505-514.

[34] Mujahid, Mahasin S., Janz, Nancy K., Hawley, Sarah T., Griggs, Jennifer J., Hamilton, Ann S., Graff, John, Katz, Steven J., (2011), “Racial/ethnic differences in job loss for women with breast cancer”, Journal of Cancer survivorship, research and practice, 5, pp. 102-111.

[35] Oberst K., Bradley C.J., Gardiner J.C., Schenk M., Given C.W. (2010), “Work task disability in employed breast and prostate cancer patients”, Journal of Cancer survivorship, research and practice, 4 (4), pp. 322-330.

[36] Paraponaris A., Sagaon Teyssier L., Ventelou B., (2010), “Job tenure and self-reported workplace discrimination for cancer survivors 2 years after diagnosis: Does employment legislation matter?”, Health Policy, 98, pp.144–155.

[37] Quinlan E., Thomas-MacLean R., Hack T., Kwan W., Miedema B., Tatemichi S., Towers A., Tilley A. (2009) "The impact of breast cancer among Canadian women: disability and productivity", Work, 34 (3), pp. 285-296.

[38] Taskila-Abrandt T., Pukkala E., Martikainen R., Karjalainen A., Hietanen P. (2005), "Employment status of Finnish cancer patients in 1997", Psycho-oncology, 14 (3), pp. 221-226.

[39] Torp S., Gudbergsson S., Dahl A., Fossa S. and Fløtten T., (2011), "Social support at work and work changes among cancer survivors in Norway" Scandinavian Journal of Public Health, 39(6), pp. 33-42.

Table 1: Sample statistics

a : age in 2008
r : first job relative revenue class

Variable	No chronic disease	Breast cancer	Difference
$a \leq 35$	32.6%	2.1%	+30.5%
$35 < a \leq 45$	27.8%	11.4%	+16.4%
$45 < a \leq 55$	19.0%	30.4%	-11.4%
$a > 55$	20.6%	56.1%	-35.5%
$r \leq Q_1$	23.9%	25.5%	-1.6%
$Q_1 < r \leq Me$	25.2%	20.1%	+5.1%
$Me < r \leq Q_3$	25.4%	23.6%	+1.8%
$r > Q_3$	25.5%	30.9%	-5.4%

Table 2: Effect of a breast cancer (C50), time related conditioning

Difference-in-differences with dynamic matching estimates. Matching variables : lagged outcome dummies (employment, unemployment, disease), year of birth (exact), first job relative revenue class (r, 4 levels), past disease class (3 levels), past unemployment class (c,3 levels).

Time after event	Treated	Matched	# twins	Employment			Without unemployment		
				Treated lagged	ATT	ASE	Treated lagged	ATT	ASE
Full sample				t_i-1			t_i-1		
t_i+1	2547	100.0%	605	0.861	-0.098*	0.005	0.765	-0.063*	0.005
t_i+2	2221	100.0%	576	0.868	-0.087*	0.005	0.771	-0.071*	0.006
t_i+3	1934	100.0%	557	0.873	-0.088*	0.006	0.777	-0.076*	0.006
t_i+4	1644	99.9%	526	0.877	-0.094*	0.006	0.785	-0.090*	0.007
t_i+5	1410	100.0%	494	0.878	-0.121*	0.007	0.783	-0.105*	0.008
Age at cancer ≤ 48				t_i-1			t_i-1		
t_i+1	1310	100.0%	684	0.897	-0.101*	0.006	0.800	-0.065*	0.006
t_i+2	1185	100.0%	644	0.900	-0.075*	0.006	0.806	-0.064*	0.007
t_i+3	1100	100.0%	617	0.899	-0.060*	0.006	0.802	-0.055*	0.007
t_i+4	986	100.0%	583	0.901	-0.057*	0.007	0.811	-0.057*	0.008
t_i+5	892	100.0%	555	0.896	-0.080*	0.008	0.805	-0.070*	0.009
Age at cancer > 48				t_i-1			t_i-1		
t_i+1	1237	100.0%	514	0.822	-0.084*	0.006	0.729	-0.052*	0.007
t_i+2	1036	100.0%	486	0.822	-0.083*	0.008	0.731	-0.062*	0.008
t_i+3	834	100.0%	464	0.830	-0.102*	0.009	0.745	-0.082*	0.010
t_i+4	658	100.0%	432	0.839	-0.116*	0.010	0.749	-0.111*	0.011
t_i+5	518	100.0%	390	0.849	-0.152*	0.011	0.749	-0.130*	0.012
Birth year ≤ 1952				t_i-1			t_i-1		
t_i+1	1344	100.0%	499	0.821	-0.081*	0.006	0.728	-0.053*	0.007
t_i+2	1203	100.0%	485	0.836	-0.092*	0.008	0.742	-0.073*	0.008
t_i+3	1055	100.0%	480	0.845	-0.106*	0.009	0.758	-0.092*	0.010
t_i+4	918	99.9%	460	0.853	-0.120*	0.010	0.767	-0.117*	0.011
t_i+5	811	100.0%	437	0.863	-0.147*	0.011	0.769	-0.125*	0.012
Birth year > 1952				t_i-1			t_i-1		
t_i+1	1203	100.0%	724	0.905	-0.115*	0.009	0.806	-0.074*	0.008
t_i+2	1018	100.0%	683	0.905	-0.082*	0.007	0.805	-0.069*	0.008
t_i+3	879	99.9%	650	0.905	-0.067*	0.007	0.801	-0.057*	0.008
t_i+4	726	100.0%	608	0.906	-0.061*	0.007	0.809	-0.056*	0.009
t_i+5	599	100.0%	571	0.898	-0.086*	0.008	0.801	-0.077*	0.010
Cancer onset ≤ 1990				t_i-1			t_i-1		
t_i+1	157	100.0%	269	0.847	-0.103*	0.011	0.783	-0.093*	0.013
t_i+2	156	100.0%	269	0.840	-0.113*	0.011	0.776	-0.084*	0.014
t_i+3	158	100.0%	267	0.842	-0.070*	0.013	0.778	-0.059*	0.015
Cancer onset 1991 – 1999				t_i-1			t_i-1		
t_i+1	810	100.0%	439	0.850	-0.086*	0.006	0.752	-0.061*	0.007
t_i+2	787	100.0%	440	0.856	-0.079*	0.007	0.759	-0.074*	0.008
t_i+3	757	100.0%	447	0.855	-0.092*	0.008	0.769	-0.088*	0.009
Cancer onset ≥ 2000				t_i-1			t_i-1		
t_i+1	1580	100.0%	659	0.868	-0.093*	0.006	0.772	-0.053*	0.006
t_i+2	1278	100.0%	618	0.880	-0.072*	0.006	0.779	-0.053*	0.007
t_i+3	1019	100.0%	585	0.892	-0.066*	0.007	0.785	-0.050*	0.007

ATT : Average effect of the treatment on the treated. ASE : Asymptotic standard error. $*$: significant at 5%. † : significant at 10%.

Table 3: Effect of a breast cancer (C50), socioeconomic conditioning

Difference-in-differences with dynamic matching estimates. Matching variables : lagged outcome dummies (employment, unemployment, disease), year of birth (exact), first job relative revenue class (r, 4 levels), past disease class (3 levels), past unemployment class (c, 3 levels).

Time after event	Treated	Matched	# twins	Employment			Without unemployment		
				Treated lagged	ATT	ASE	Treated lagged	ATT	ASE
$r \leq Q_1$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	615	100,0%	563	0,795	-0,152*	0,012	0,676	-0,101*	0,013
$t_i + 2$	534	100,0%	533	0,799	-0,132*	0,013	0,677	-0,107*	0,013
$t_i + 3$	465	100,0%	513	0,804	-0,137*	0,014	0,685	-0,126*	0,015
$t_i + 4$	381	100,0%	488	0,811	-0,122*	0,015	0,690	-0,110*	0,016
$t_i + 5$	330	100,0%	459	0,821	-0,176*	0,017	0,694	-0,165*	0,018
$Q_1 < r \leq Me$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	499	100,0%	509	0,840	-0,119*	0,010	0,697	-0,071*	0,011
$t_i + 2$	440	100,0%	479	0,848	-0,097*	0,011	0,702	-0,075*	0,013
$t_i + 3$	377	100,0%	456	0,859	-0,081*	0,012	0,713	-0,063*	0,015
$t_i + 4$	329	100,0%	412	0,857	-0,079*	0,014	0,705	-0,067*	0,016
$t_i + 5$	289	100,0%	381	0,848	-0,083*	0,014	0,685	-0,077*	0,017
$Me < r \leq Q_3$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	649	100,0%	653	0,892	-0,093*	0,010	0,807	-0,060*	0,010
$t_i + 2$	567	100,0%	621	0,899	-0,082*	0,010	0,811	-0,060*	0,011
$t_i + 3$	488	100,0%	600	0,902	-0,081*	0,011	0,807	-0,066*	0,012
$t_i + 4$	406	99,8%	558	0,901	-0,082*	0,013	0,820	-0,089*	0,014
$t_i + 5$	341	100,0%	517	0,903	-0,103*	0,013	0,827	-0,084*	0,015
$r > Q_3$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	784	100,0%	660	0,899	-0,047*	0,007	0,843	-0,031*	0,008
$t_i + 2$	680	100,0%	635	0,907	-0,051*	0,008	0,854	-0,049*	0,010
$t_i + 3$	604	100,0%	620	0,911	-0,061*	0,010	0,864	-0,054*	0,011
$t_i + 4$	528	100,0%	598	0,917	-0,092*	0,011	0,877	-0,090*	0,013
$t_i + 5$	450	100,0%	574	0,920	-0,119*	0,014	0,878	-0,095*	0,015
$c = 0$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	1217	100,0%	351	0,917	-0,054*	0,007	0,917	-0,047*	0,008
$t_i + 2$	1082	100,0%	352	0,920	-0,057*	0,008	0,920	-0,063*	0,009
$t_i + 3$	967	100,0%	352	0,918	-0,080*	0,009	0,918	-0,077*	0,010
$t_i + 4$	848	99,9%	348	0,920	-0,089*	0,010	0,920	-0,099*	0,011
$t_i + 5$	739	100,0%	342	0,924	-0,120*	0,012	0,924	-0,110*	0,012
$0 < c \leq 0.16$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	743	100,0%	163	0,846	-0,114*	0,008	0,704	-0,077*	0,009
$t_i + 2$	634	100,0%	149	0,853	-0,102*	0,009	0,710	-0,079*	0,010
$t_i + 3$	533	100,0%	142	0,862	-0,099*	0,010	0,720	-0,096*	0,011
$t_i + 4$	446	100,0%	130	0,871	-0,109*	0,010	0,727	-0,109*	0,012
$t_i + 5$	367	100,0%	114	0,874	-0,125*	0,011	0,717	-0,111*	0,012
$c > 0.16$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	587	100,0%	283	0,764	-0,166*	0,012	0,533	-0,079*	0,012
$t_i + 2$	505	100,0%	261	0,777	-0,127*	0,012	0,531	-0,082*	0,013
$t_i + 3$	434	100,0%	247	0,789	-0,093*	0,011	0,539	-0,053*	0,014
$t_i + 4$	350	100,0%	229	0,787	-0,083*	0,013	0,543	-0,043*	0,015
$t_i + 5$	304	100,0%	215	0,781	-0,119*	0,013	0,530	-0,087*	0,016

ATT : Average effect of the treatment on the treated. ASE : Asymptotic standard error. $*$: significant at 5%. † : significant at 10%.

A Estimation of the standard errors

This section presents the computation of the standard errors without using the bootstrap, in order to save computing time. The ATT estimator, denoted $\hat{c}$, can be written :

$$\hat{c} = \frac{1}{I}\sum_{i\in I}\hat{c}_i \text{ with } \hat{c}_i = \Delta y_i - \frac{1}{J(i)}\sum_{j\in J(i)}\Delta y_j$$

where I is the treated set (and their number), y_i is the outcome variable of individual i, Δy_i is the before-after difference and $J(i)$ is the set of individual i's twins (and their number). The previous formula defines the difference in differences estimator. If two treated individuals have the same matching variables, and if their treatment happens at the same date, they will be matched with exactly the same twins so that the same mean will be subtracted from their outcome variable. We regroup the treated according to their matching variables and treatment date. Let $k \in K$ be a specific vector regrouping the matching variables and the treatment date, the set of all the treated individuals in the matching group k (and their number) is defined by:

$$I(k) = \{i \in I : (X_i, t_i) = k\},\ k \in K$$

and we let $J(k)$ denote the common twin's set of the treated in group k (and their number). By definition, the $I(k)$ sets define a partition of the treated set $I = \bigcap_k I(k), I(k)\bigcap I(k') = \emptyset \forall k \neq k'$. Therefore, the ATT can be rewritten:

$$\hat{c} = \frac{1}{I}\sum_{k\in K}\sum_{i\in I(k)}\left(\Delta y_i - m_k\right)$$

where $m_k = J(k)^{-1}\sum_{j\in J(k)}\Delta y_j$ is the twin's common mean inside group k. Simplifying inside the sum, we get :

$$\hat{c} = \frac{1}{I}\sum_{k\in K}\left\{\sum_{i\in I(k)}\Delta y_i - I(k)m_k\right\}$$

In order to compute the variance of this estimator, we make the standard independence assumption between the y_i's. First, we notice that the groups k are independent of each other since they have neither a treated nor a twin in common. We get :

$$\mathrm{V}(\hat{c}) = \frac{1}{I^2}\sum_{k\in K}\mathrm{V}\left(\sum_{i\in I(k)}\Delta y_i - I(k)m_k\right)$$

Second, we notice that the y_i's are independent from the m_k's since they are computed from different individuals. We get :

$$\begin{aligned}\mathrm{V}(\hat{c}) &= \frac{1}{I^2}\sum_{k\in K}\left\{\mathrm{V}\left(\sum_{i\in I(k)}\Delta y_i\right) + I(k)^2\mathrm{V}(m_k)\right\}\\ &= \sum_{k\in K}\left(\frac{I(k)}{I}\right)^2\left\{\mathrm{V}\left(m_k^T\right) + \mathrm{V}(m_k)\right\}\end{aligned}$$

with $m_k^T = I(k)^{-1}\sum_{i\in I(k)}\Delta y_i$ the mean outcome of the treated inside group k. The estimator is obtained by replacing the theoretical statistics by their empirical counterparts.

Table 4: Effect of a breast cancer, complementary variables (C50), time related conditioning

Difference-in-differences with dynamic matching estimates. Matching variables : lagged outcome dummies (employment, unemployment, disease), year of birth (exact), first job relative revenue class (r, 4 levels), past disease class (3 levels), past unemployment class (c,3 levels).

Time after event	Treated	Matched	# twins	Disease			Unemployment		
				Treated lagged	ATT	ASE	Treated lagged	ATT	ASE
Full sample				t_i-1			t_i-1		
t_i+1	2547	100,0%	605	0,056	0,528*	0,007	0,163	-0,037*	0,004
t_i+2	2221	100,0%	576	0,057	0,274*	0,007	0,160	-0,007	0,005
t_i+3	1934	100,0%	557	0,058	0,147*	0,007	0,152	0,007	0,006
t_i+4	1644	99,9%	526	0,058	0,047*	0,005	0,143	0,015*	0,006
t_i+5	1410	100,0%	494	0,059	0,049*	0,006	0,144	0,012†	0,007
Age at cancer ≤ 48				t_i-1			t_i-1		
t_i+1	1310	100,0%	684	0,072	0,548*	0,009	0,140	-0,032*	0,005
t_i+2	1185	100,0%	644	0,070	0,268*	0,009	0,138	0,006	0,006
t_i+3	1100	100,0%	617	0,072	0,140*	0,008	0,139	0,017*	0,006
t_i+4	986	100,0%	583	0,070	0,050*	0,006	0,131	0,019*	0,007
t_i+5	892	100,0%	555	0,070	0,056*	0,007	0,132	0,003	0,007
Age at cancer > 48				t_i-1			t_i-1		
t_i+1	1237	100,0%	514	0,039	0,486*	0,009	0,188	-0,042*	0,005
t_i+2	1036	100,0%	486	0,039	0,256*	0,009	0,188	-0,025*	0,007
t_i+3	834	100,0%	464	0,042	0,123*	0,009	0,186	-0,010	0,009
t_i+4	658	100,0%	432	0,040	0,007	0,007	0,168	0,005	0,010
t_i+5	518	100,0%	390	0,039	-0,005	0,007	0,162	0,023†	0,012
Birth year ≤ 1952				t_i-1			t_i-1		
t_i+1	1344	100,0%	499	0,044	0,469*	0,010	0,184	-0,033*	0,005
t_i+2	1203	100,0%	485	0,044	0,254*	0,010	0,173	-0,010	0,007
t_i+3	1055	100,0%	480	0,046	0,133*	0,009	0,155	0,008	0,009
t_i+4	918	99,9%	460	0,046	0,037*	0,007	0,146	0,019†	0,010
t_i+5	811	100,0%	437	0,047	0,034*	0,008	0,147	0,020*	0,010
Birth year > 1952				t_i-1			t_i-1		
t_i+1	1203	100,0%	724	0,070	0,594*	0,010	0,140	-0,041*	0,005
t_i+2	1018	100,0%	683	0,073	0,298*	0,010	0,145	-0,002	0,007
t_i+3	879	99,9%	650	0,073	0,163*	0,010	0,148	0,005	0,007
t_i+4	726	100,0%	608	0,074	0,060*	0,007	0,139	0,011	0,008
t_i+5	599	100,0%	571	0,075	0,070*	0,008	0,140	0,000	0,008
Cancer onset ≤ 1990				t_i-1			t_i-1		
t_i+1	157	100.00%	269	0.076	0.381*	0.022	0.096	0.024*	0.008
t_i+2	156	100.00%	269	0.071	0.170*	0.019	0.096	0.015	0.011
t_i+3	158	100.00%	267	0.076	0.127*	0.018	0.095	-0.007	0.009
Cancer onset 1991 – 1999				t_i-1			t_i-1		
t_i+1	810	100.00%	439	0.048	0.466*	0.010	0.159	-0.034*	0.006
t_i+2	787	100.00%	440	0.047	0.235*	0.010	0.157	0.007	0.008
t_i+3	757	100.00%	447	0.048	0.131*	0.009	0.141	0.030*	0.008
Cancer onset ≥ 2000				t_i-1			t_i-1		
t_i+1	1580	100.00%	659	0.057	0.559*	0.008	0.171	-0.046*	0.004
t_i+2	1278	100.00%	618	0.061	0.292*	0.009	0.170	-0.022*	0.006
t_i+3	1019	100.00%	585	0.061	0.137*	0.008	0.168	-0.014*	0.006

ATT : Average effect of the treatment on the treated. ASE : Asymptotic standard error. $*$: significant at 5%. † : significant at 10%.

Table 5: Effect of a breast cancer (C50), complementary variables, socioeconomic conditioning

Difference-in-differences with dynamic matching estimates. Matching variables : lagged outcome dummies (employment, unemployment, disease), year of birth (exact), first job relative revenue class (r, 4 levels), past disease class (3 levels), past unemployment class (c, 3 levels).

Time after event	Treated	Matched	# twins	Disease			Unemployment		
				Treated lagged	ATT	ASE	Treated lagged	ATT	ASE
$r \leq Q_1$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	615	100,0%	563	0,039	0,510*	0,014	0,208	-0,048*	0,010
$t_i + 2$	534	100,0%	533	0,039	0,295*	0,014	0,212	-0,013	0,011
$t_i + 3$	465	100,0%	513	0,039	0,151*	0,013	0,192	0,017	0,013
$t_i + 4$	381	100,0%	488	0,037	0,044*	0,010	0,189	0,011	0,014
$t_i + 5$	330	100,0%	459	0,030	0,029*	0,010	0,185	0,043*	0,015
$Q_1 < r \leq Me$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	499	100,0%	509	0,074	0,486*	0,015	0,214	-0,043*	0,008
$t_i + 2$	440	100,0%	479	0,073	0,215*	0,014	0,209	0,002	0,011
$t_i + 3$	377	100,0%	456	0,074	0,122*	0,013	0,202	0,016	0,014
$t_i + 4$	329	100,0%	412	0,076	0,035*	0,011	0,210	0,029*	0,014
$t_i + 5$	289	100,0%	381	0,076	0,060*	0,011	0,221	0,009	0,014
$Me < r \leq Q_3$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	649	100,0%	653	0,052	0,588*	0,014	0,134	-0,039*	0,007
$t_i + 2$	567	100,0%	621	0,060	0,323*	0,015	0,136	-0,013	0,009
$t_i + 3$	488	100,0%	600	0,061	0,172*	0,014	0,139	0,012	0,011
$t_i + 4$	406	99,8%	558	0,064	0,045*	0,010	0,123	0,025*	0,012
$t_i + 5$	341	100,0%	517	0,067	0,046*	0,011	0,120	0,004	0,013
$r > Q_3$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	784	100,0%	660	0,061	0,518*	0,014	0,120	-0,023*	0,006
$t_i + 2$	680	100,0%	635	0,059	0,256*	0,014	0,109	-0,002	0,009
$t_i + 3$	604	100,0%	620	0,060	0,139*	0,012	0,099	-0,011	0,009
$t_i + 4$	528	100,0%	598	0,059	0,058*	0,010	0,083	0,003	0,011
$t_i + 5$	450	100,0%	574	0,062	0,059*	0,012	0,082	-0,004	0,013
$c = 0$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	1217	100,0%	351	0,044	0,526*	0,012	0,000	-0,008	0,005
$t_i + 2$	1082	100,0%	352	0,047	0,277*	0,012	0,000	0,015*	0,007
$t_i + 3$	967	100,0%	352	0,050	0,143*	0,011	0,000	0,022*	0,008
$t_i + 4$	848	99,9%	348	0,050	0,051*	0,008	0,000	0,035*	0,010
$t_i + 5$	739	100,0%	342	0,050	0,056*	0,009	0,000	0,035*	0,010
$0 < c \leq 0.16$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	743	100,0%	163	0,057	0,514*	0,011	0,239	-0,034*	0,007
$t_i + 2$	634	100,0%	149	0,055	0,249*	0,010	0,230	-0,013	0,008
$t_i + 3$	533	100,0%	142	0,051	0,135*	0,010	0,214	0,007	0,010
$t_i + 4$	446	100,0%	130	0,050	0,031*	0,008	0,205	0,019†	0,011
$t_i + 5$	367	100,0%	114	0,052	0,035*	0,008	0,206	0,009	0,011
$c > 0.16$				$t_i - 1$			$t_i - 1$		
$t_i + 1$	587	100,0%	283	0,075	0,551*	0,013	0,406	-0,099*	0,010
$t_i + 2$	505	100,0%	261	0,074	0,307*	0,014	0,417	-0,044*	0,013
$t_i + 3$	434	100,0%	247	0,076	0,170*	0,013	0,414	-0,030*	0,014
$t_i + 4$	350	100,0%	229	0,080	0,056*	0,010	0,408	-0,037*	0,014
$t_i + 5$	304	100,0%	215	0,076	0,047*	0,010	0,417	-0,044*	0,015

ATT : Average effect of the treatment on the treated. ASE : Asymptotic standard error. $*$: significant at 5%. † : significant at 10%.