Document de travail du LEM / Discussion paper LEM 2019-06

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

Thomas BARNAY

ERUDITE, Université Paris Est-Créteil, IST-PE and TEPP (FR CNRS 3435)

Emmanuel DUGUET

ERUDITE, Université Paris Est-Créteil, IST-PE, TEPP (FR CNRS 3435) and CEET

Christine LE CLAINCHE

LEM UMR 9221 / christine.leclainche@gmail.com

http://lem.cnrs.fr/

The effects of breast cancer on individual labour market outcomes: an evaluation from an administrative panel in France[1]

Thomas BARNAY
ERUDITE, Université Paris Est-Créteil, IST-PE and TEPP (FR CNRS 3435)

Emmanuel DUGUET
ERUDITE, Université Paris Est-Créteil, IST-PE, TEPP (FR CNRS 3435) and CEET

Christine LE CLAINCHE
Université de Lille, LEM (UMR CNRS 9221) and CEET

April 2019

Abstract. We estimate the effect of breast cancer on labour market participation in France, up to five years after the cancer onset. The causal inference is systematically confounded by differences in age, wage at the beginning of career, past health and the past history in the labour market. We account for all these issues thanks to an administrative data set which follows individuals from their entry in the labour market. We find that the detrimental effect of the breast cancer for women increases significantly over time up to 10 percentage points after five years. We also find evidence that the effect of cancer is reduced for younger generations.

JEL: I10, J21, J22.

Keywords: breast cancer, female labour market participation, difference in differences, matching, France.

[1] Financial support from the French National Cancer Institute (INCA) and Fondation ARC is gratefully acknowledged. We thank the participants to the following seminars and conferences for their helpful suggestions and comments: 37th JESF conference (Dijon, 2015), GAINS seminar (Le Mans, 2016), 13th TEPP Conference (La Réunion, 2016), 4th IAAE Conference (Sapporo, 2017).

1. Introduction

Thanks to advances in organized screening, detection and treatment, cancers may be related to chronic diseases (Cutler, 2008). Breast cancer produces a substantial burden on the statutory health insurance in France. This cost burden affects the long-term disease scheme, which supports all expenditures related to chronic diseases, including cancers.[2] In the female population, breast cancer is also the most prevalent and the first cause of mortality by cancer. It exhibits an earlier onset than that of most other cancers[3] and requires treatments associated with functional sequels and therefore raises questions regarding the impact of breast cancer on individual well-being and, especially, labour market outcomes. Furthermore, the net survival rate of women diagnosed with breast cancer in France between 1989 and 2004 is 97% after one year and 86% after five years (Jooste, Grosclaude, Remontet, Launoy, Baldi, Molinie, Arveux, Bossard, Bouvier and Colonna, 2013). This is one of the highest rate of return to work among female cancers. Except Thyroid cancer, all the cancers affecting women are associated to a rather shorter net survival rate at 5 or 10 years (Inca, 2019).

The effect of cancer on labour market outcomes.

According to the health capital model of Grossman (1972), health is considered as durable capital good which depreciates over time. Thus, health appears both as a consumption and an investment good. According to this theoretical point of view, the onset of cancer, like any serious health event, affects career paths through the potential effects on the health stock, the decrease in productivity and in hours worked (Moran, Short and Hollenbeak, 2011), the depreciation rate of health capital and the future investments in human capital.

By and large, the negative impact of cancer on the career path operates primarily through functional limitations (Bradley et al. (2002) in the USA), which may be specific, such as arm pain for breast cancer, as a major sequel of treatment (Quinlan, Thomas-MacLean, Hack, Kwan, Miedema, Tatemichi, Towers and Tilley, 2009, in Canada; Blinder, Patil, Thind, Diamant, Hudis, Basch and Maly, 2012, in the USA), depressive episodes (Damkjaer, Deltour, Suppli, Christensen, Kroman, Johansen and Dalton, 2011, in Denmark) and memory and concentration disorders (Oberst, Bradley, Gardiner, Schenk and Given, 2010, in the 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, the type of cancer (site, severity of the disease) and, finally, the nature of the treatment and sequelae (Johnsson, Fornander, Rutqvist and Olsson, 2011; Lindbohm, Kuosma, Taskila, Hietanen, Carlsen, Gudbergsson

[2] This mechanism is known as ALD in France (ALD for *Affections de Longue Durée*).

[3] The median age at breast cancer onset was approximately 67 years in France in 2018 (INCA, 2019). Cancers before 40 years old represent 5% of the cases diagnosed. Diagnoses of new cases are generally made after 55, partly due to systematic screening from 50 years onward (INCA, 2014).

and Gunnarsdottir, 2011; Mujahid, Janz, Hawley, Griggs, Hamilton, Graff and Katz, 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, predicts the occurrence of occupation-related cancers. Feuerstein, Todd, Moskowitz, Bruns, Stoler, Nassif and Yu (2010) stress the importance of improvements in the workplace in terms of schedule flexibility, social support from colleagues, social climate and job stress to protect working ability among cancer survivors. From a linkage of different Canadian data, Jeon (2017) studies the heterogeneity of the cancer effect by considering different subpopulations by age, education, and pre-diagnosis earnings. The probability of working and the annual earnings are significantly higher for the cancer survivors with a high school diploma and the older survivors. The literature also shows that the labour market outcomes of spouses may be affected after one spouse's cancer diagnosis (Jeon and Pohl, 2017).

Focus on Breast cancer.

A large international literature (with a particular reliance on US data) is specifically devoted to the effect of breast cancer on professional paths (Chirikos, Russell-Jacobs and Cantor 2002; Chirikos, Russell-Jacobs and Jacobsen 2002; Drolet, Maunsell, Mondor, Brisson, Brisson, Masse and Deschenes 2005; Bradley, Oberst and Schenk, 2006; Bradley, Neumark and Barkowski, 2013; Heinesen and Kolodziejczyk, 2013 and 2016).

For instance, Bradley, Neumark, Luo and Bednarek (2007) show that the negative effect of cancer on employment significantly persists 6 months after the diagnosis but not more. Heinesen and Kolodziejczyk (2013) measure the 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) using propensity score methods with the persons without cancer as a control group. From Danish data (2001-2009), Carlsen, Badsberg and Dalton (2014) stress that women, after a breast cancer diagnosis, who experienced periods of unemployment before the diagnosis have an increased risk of being unemployed thereafter relative to women who worked before. Past french studies are more limited. Eichenbaum-Voline, Malavolti, Paraponaris and Ventelou (2008) and Joutard, Paraponaris, Sagaon-Teyssier and Ventelou (2012) apply a matching method to survey data that include treatment variables. Marino, Sagaon, Malavolti and Le Coroller- Soriano (2013) show that, two years after the diagnosis of cancer, the probability of returning to work in the female population is 72%.

Many articles underline the role of different health and socioeconomic characteristics that influence the effect of breast cancer on employment. First, a significant body of literature highlights the nature

of cancer and types of treatment (Hassett, O'Malley and Keating, 2009; Jagsi, Hawley, Abrahamse, Li, Janz, Griggs, Bradley, Graff, Hamilton and Kratz, 2014). 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, facing a combination of treatments (especially chemotherapy before and/or after radiotherapy) has the most detrimental effect on the speed with which a woman with breast cancer can return to work (for a french data set that includes the severity of the disease and the type of treatments, see Duguet and Le Clainche, 2016). Women who have undergone a surgery with a partial mastectomy followed by radiotherapy can often return to work in the 6 months following the surgery, if no comorbidity occurs.

Treatment and comorbidity factors, sociodemographic and work-related characteristics may also explain differences in labour outcomes (Bradley, Neumark, Bednarek and Schenk, 2004; Torp, Gudbergsson, Dahl, Fossa and Fløtten, 2011). Cross-country differences in the delays beyond 6 months may be explained by protective factors such as favourable social protection rules (sick leave legislation, social insurance schemes, work flexibility and adjustments) and how employers are allowed to arrange working conditions. Using data, Duguet and Le Clainche (2016) show that the probability of returning to work two years after diagnosis, especially for women diagnosed with breast cancer, increases when appropriate working conditions adjustments are implemented. Generally, the onset of cancer affects future investments in human capital (primary or secondary health prevention) due to the difficulty of combining work and cancer treatment (Yarker et al., 2010; Johnsson et al., 2011).

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 (Farley, Vasey and Moran (2008), Paraponaris, Teyssier and Ventelou (2010), Petersson et al. (2011), Torp et al. (2011)).

From Swedish data, Petersson et al. (2011) find that, one month after the surgery, 56% of women with breast cancer are on sick leave, the majority full-time. According to Farley, Vasey and Moran (2008), in the USA, survivorship affected the probability of working full-time and hours worked for both genders 2-6 years post-diagnosis. Torp et al. (2011), using a Norwegian database, highlight that a low socioeconomic position appears to be a risk factor for returning to work. On the basis of Korean data (1993-2002), after a breast cancer diagnosis, working women are more frequently unemployed if they have low education or a low income (Eunmi, Cho, Shin, Park, Ahn, Noh, Nam, Lee and Yun, 2009). , Kolodziejczyk and Heinesen (2016), on Danish data, stress out differences between the public and private sector employees, in the effect of breast cancer on the probability of being out of the labour

force three years after the diagnosis. Their findings underline a more pronounced low educated adverse effect in the public sector.

Using french data, Paraponaris, Teyssier and Ventelou (2010) study the relationship between cancer occurrence and the type of labour contract. Their findings indicate that fixed-term contracts exhibit greater risk of job loss for workers in the female population (-8 pp relative to permanent contracts).

From a theoretical perspective, the return to work depends on economic incentives. Bradley, Neumark and Barkowski (2013) demonstrate that the negative effect of breast cancer on employment is reduced if the patient's health insurance is dependent on the job. This result refers to the "job lock" assumption, i.e., workers remain in their current job to maintain their health insurance. In contrast to the USA, in France, for particular diseases that require intensive, expensive and long-term care (such as cancer), the long-term disease scheme makes the health insurance independent from the job. In this study, we focus on the sociodemographic and work-related aspects of cancer. Given the french legal framework, we cannot assume a job lock effect.

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 on employment outcomes up to five years after its onset. We perform a difference-in-differences analysis combined with a dynamic matching algorithm. Second, we highlight the role of protective factors that attenuate the adverse effects of cancer on labour market outcomes. Thanks to our data, we can examine four potential types of socioeconomic factors that could influence employment in the short and long term. First, breast cancer occurring at a younger age could be less disadvantageous than at a later age (Petersson, 2011). Second, we account for the stability of the past career before the onset of cancer (Heckman and Borjas, 1980; Gregg and Tominey, 2005). Third, we account for the health history of the workers. Finally and more originally, we also examine a technological effect by considering a large period of diagnosis as contrary to literature. Advances in medicine could lead to a better return to work for the most recent generation of cancer survivors.

The paper is organized as follows. Section 2 presents the data and section 3 discusses the model and the estimation method. The results are reported and discussed in section 4.

2. Data

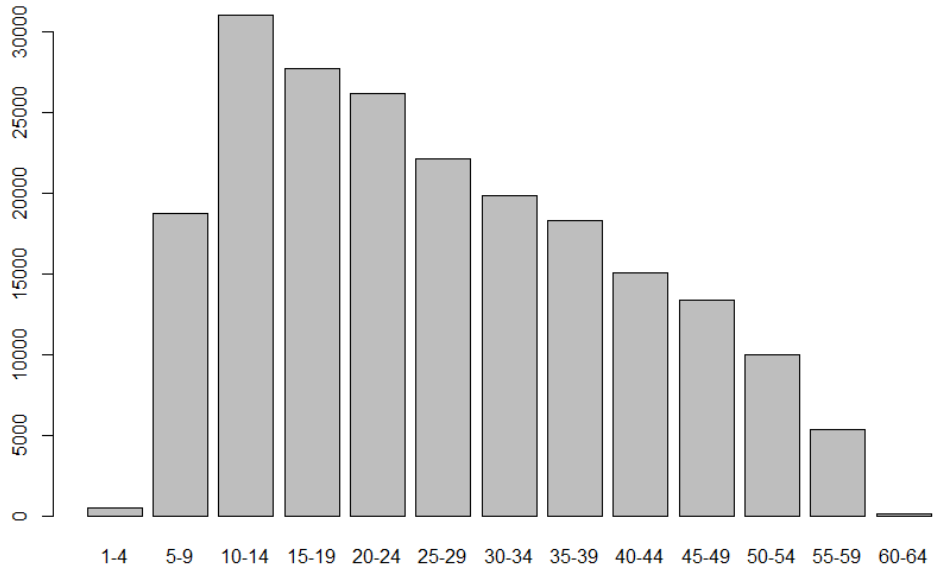
Source. We use the HYGIE data set, which was constructed from two nationwide administrative sources: the National Pension Fund (CNAV) and the National Health Insurance Fund for Salaried Workers (CNAM). The resulting sample contains individual information on the recipients, their

professional careers, medical consumption and sick leaves. HYGIE is representative of the private sector workers in France.[4]

The files were extracted from the National Career Management System (CNAV-SNGC), which gathers information about all private sector employees in France, and the National Recipients Statistical System (CNAV-SNSP), which includes all private sector retired workers in France. These data are matched with the sickness benefit data taken from the National Health Insurance Inter-regime Information System (CNAM-SNIIR-AM). We obtain a random sample of the recipients aged 22 to 70 years who contributed to the general pension fund at least once in their live. The CNAM data provides information about the recipients of the National Health Insurance scheme, who received sickness benefits for at least one health service in 2003, 2004 or 2005. The merger of the CNAV and CNAM data enabled us to construct the HYGIE database panel. The total sample includes information about 552,048 workers, including 225,331 women. These women are followed every year since their entry in the labour market, during 23 years on average, leaving us with 5,161,332 observations (Figure 1).

Figure 1: Number of years of data per worker

Reading example: About 15,000 workers have between 40 and 44 years of data.



[4] Hygie is not an acronym. In the Greek as well as Roman mythology, Hygieia (Hygie in French) is a godess of health. Her name is the source of the work hygiene.

Table 1. Attrition

t_i : Cancer onset date. Reading example: if we consider the differences between $t_i - 1$ and $t_i + 3$, 94.9% of the differences were missing at random and 5.1% were missing because the patient died during this time interval.

Difference	Missing at random	Death
$t_i - 1, t_i + 1$	97.0%	3.0%
$t_i - 1, t_i + 2$	94.6%	5.4%
$t_i - 1, t_i + 3$	94.9%	5.1%
$t_i - 1, t_i + 4$	94.7%	5.3%
$t_i - 1, t_i + 5$	94.7%	5.3%

The administrative data include a sick leave dummy and the International Classification of Diseases (henceforth, ICD). We identify breast cancer with the ICD code C50, on the basis of a first registration in the long-term disease scheme (so called ALD in french).

The french scheme for long term diseases is very specific and without equivalent elsewhere in Europe. It offers full free care at the point of services but only for related cancer cares. It concerns all the cancers and diseases which are potentially long and costly, like severe heart diseases, diabetes or kidney insufficiency. The recognition of the ALD scheme is done by the gatekeeper (generally a General Practitioner) in connection with the referent oncologist who is in charge of the patient follow up. The access to this recognition is strictly controlled by the social security.

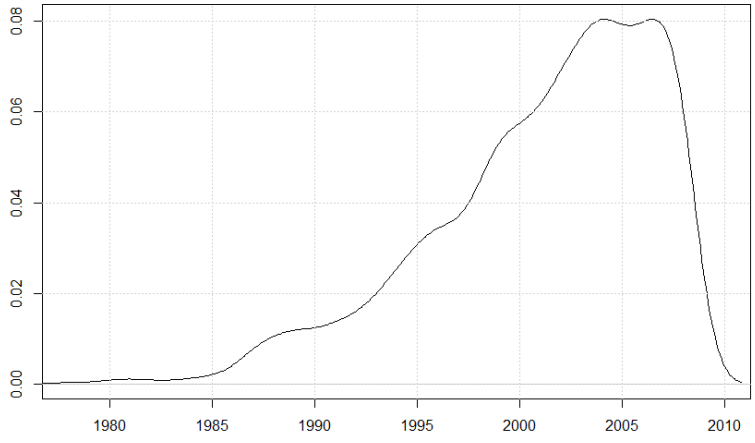
However, the situation of the workers may also depend on their status at the onset date of the cancer. The executives from big companies benefit from additional advantages regarding the duration of the sick leave and the rate of replacement. It is less the case for the employees of the small companies. In addition, the people who exit the labour market due to the disease can also benefit from disability allowances schemes, when the resources of the household they belong to are low and if they cannot return to work after the care (see Marino et al. (2013).

Since the women in the sample are rather young, we have a small attrition related to death (Table 1). In order to implement a dynamic matching methodology, we retain in the control group all the women with no long term disease yet. Therefore, the control group includes all the women with neither a cancer nor a long-term disease, as recognized as an ALD, at the time of the cancer onset but potentially suffering from other health problems. We have checked that this choice did not significantly alter our results compared to the more standard control group which excludes all cancers but includes the other

types of disease, like in Heinesen and Kolodziejczyk (2013).[5] This can be explained by the fact that the women in our control group can have diseases which are not recognized as ALD.

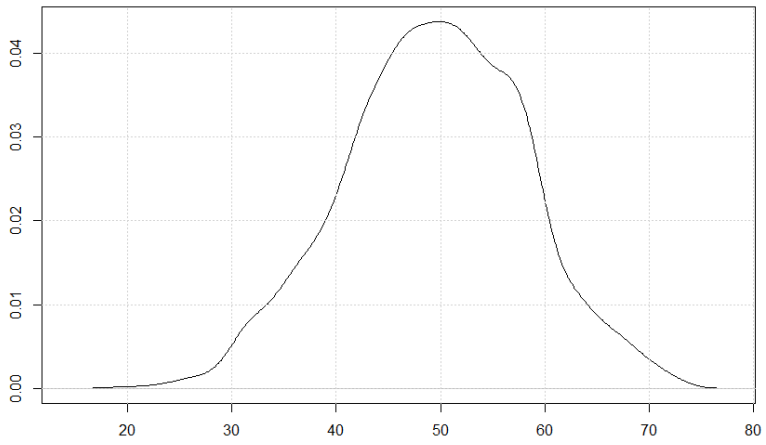
We can observe labour market participation before and after the cancer onset, whatever the date at which the cancer occurs. Figure 1 shows that the number of observations available for each woman is rather large, with an average 23 years of data. This will allow us for testing two additional hypotheses: first, does the impact of cancer vary with the date of cancer onset? Figure 2 shows the distribution of these dates. We have enough observations for studying three intervals: before 1990, 1991-1999 and 2000-2008. We expect that the progress of medicine reduces the detrimental effects of breast cancer over time. Secondly, does the effect of cancer vary with the age of the patient at the cancer onset? The median age is 48 (Figure 3), which is young for a cancer patient since the median age at diagnosis is 63 for the whole population in France (National Institute of Cancer, 2017). We will examine whether a young age is an advantage or not.

Figure 2: Breast Cancer Onset Date



[5] The results are reported in Table A.2.

Figure 3: Age at Cancer Onset



Sample. Our method uses matching. We need variables which influences both the probability of breast cancer and the participation in the labour market (our outcome variable). We will first use the year of birth, since the probability of cancer increases with age. Labour market participation also varies with age. The participation first increases at the beginning of career, reaches a maximum and decreases when the workers get closer to retirement. Secondly, we need a proxy for the education level. A low education level is associated with both a worse health and a weaker participation in the labour market. Since this information is not available in the data set, we have computed an education proxy in the following way. It relies on the fact that, especially in the french case, education is closely related to the starting wage, defined as the wage of the first full year in the labour market. By and large, the education level is associated with a better labour market situation. This positive correlation has been checked at the beginning of working life (Degorre et al., 2009 ; Le Rhun et Pollet, 2015).

We divide the starting wage of each individual by the median starting wage of the same year in order to correct for inflation, and we take four equal sized classes. Taking this ratio provides both an inflation correction, which is needed over such a long period, and a quantile interpretation. A ratio equal to 1 indicates the median, since the starting wage is expressed as a fraction of the median wage of the year of entry in the labour market. Thirdly, two other time-varying variables may influence the current participation in the labour market: past health problems could indicate a lower productivity and past participation in the labour market is related to the path dependence issue (Krueger et al., 2014). We wish to compare workers with similar health and participation histories. The health history indicator is defined as the ratio of the number of years with at least one sick leave quarter divided by the number of years in the labour market. This indicator indicates the individual proportion of years with a significant health problem in the past career of the worker. The work history indicator is defined as the

number of years with a stable employment status (see below) divided by the number of years spent in the labour market. It indicates the individual proportion of years with 4 quarters of employment contributions and no unemployment quarter. It should proxy the labour insertion quality of the worker.

The outcome variables indicating the individual's employment status were identified in the HYGIE database on the basis of the compulsory contributions paid by the workers either for their health insurance or for their retirement plan. More precisely, we measure activity through the compulsory health insurance contributions of the workers. They contribute both when they work and when they are unemployed, the difference being recorded in the data set. Each worker can contribute up to four quarters each year and we build the activity measurement through both the number of quarters and the nature of the quarters (employed/unemployed). We were able to construct four mutually exclusive statuses from HYGIE: stable employment, unstable employment, unemployment and retirement. A stable employment situation happens when a worker validates four quarters in employment. An unstable employment situation occurs when a worker validates less than four quarters in employment or when it validates both unemployment and employment quarters during the same year. An unemployment situation is defined as the validation of unemployment quarters only, whatever their number. The retirement situation is defined directly from the retirement date, which is available in the sample. Its definition excludes both employment and unemployment records. This gives four mutually exclusive dummy variables. We also use a total employment dummy, equal to the sum of the stable and unstable employment dummies. Stable employment is used to build the labour market history indicator used in the matching process.

Table 2 – Sample statistics

Variables	With breast cancer	With no chronic disease	Difference
Matching variables			
Starting relative wage (r):			
$r \leq \mathrm{Q}_1$	24.4%	23.7%	+0.7%^{ns}
$\mathrm{Q}_1 < r \leq \mathrm{Me}$	19.5%	24.9%	-5.4%**
$\mathrm{Me} < r \leq \mathrm{Q}_3$	24.4%	25.4%	-1.0%^{ns}
$r > \mathrm{Q}_3$	31.6%	25.9%	+5.7%**
Age*	48.1	37.9	+10.2**
Past employment stability (c)*:			
$c \leq 0.5$	38.1%	64.3%	-26.2%**
$0.5 < c \leq 0.7$	31.3%	20.2%	+11.1%**
$c > 0.7$	30.6%	15.5%	+15.2%**
Past health problems (h)*:			
$h = 0$	32.0%	45.2%	-13.3%**
$0 < h \leq 0.06$	36.6%	22.5%	+14.1%**
$h > 0.06$	31.4%	32.3%	-0.8%^{ns}
Outcome variables in 2008			
Employment :	48.6%	74.7%	-26.1%**
Stable employment	41.0%	62.1%	-21.1%**
Unstable employment	7.5%	12.6%	-5.1%**

* one year before the breast cancer onset. ** significant at the 1% level. ns: not significant at the 10% level.

Table 2 reports sample statistics. Women with a breast cancer appear more often both in the highest and the lowest starting wage classes. The other variables are taken one year before the cancer onset.[6] On average, the women with a breast cancer are older (+10.2 years) than the women in the control group, had a more stable employment status and more health problems in the past. These confounding variables do not all push in the same direction and an econometric analysis is clearly needed to disentangle their effect from the effect of breast cancer. The outcome variables in 2008 clearly show that women with a breast cancer work much less often than the other women (-26%) but this may be related both to breast cancer and to other factors, like age or a worse health history unrelated to cancer.

[6] We have checked that the results are similar to the ones obtained with the variables taken two years before the cancer onset.

Figure 4: Employment rate and Breast Cancer Onset

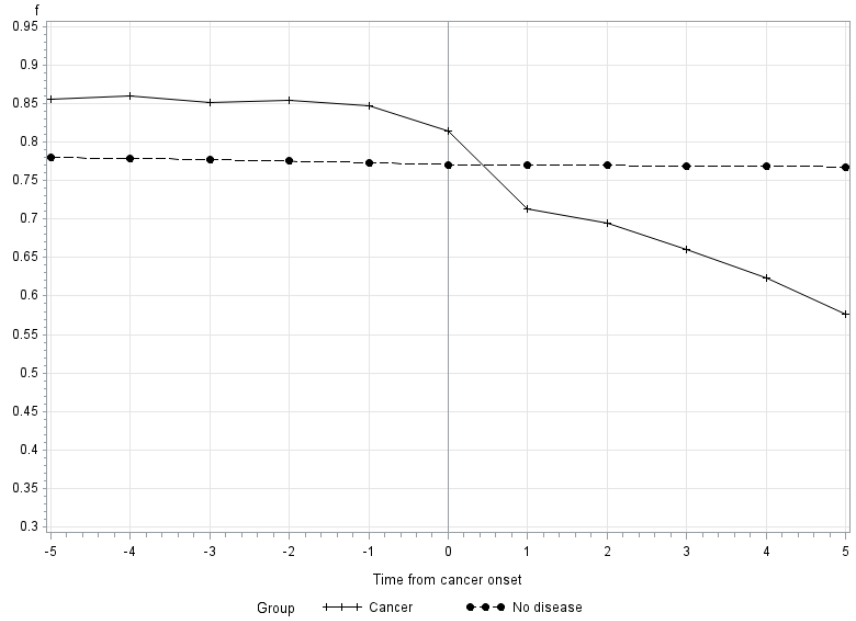


Figure 4 presents the plot of the employment participation rate against the distance from the cancer onset date. A negative number indicates a number of years before the cancer onset; a positive number indicates a number of years after the cancer onset. The doted curve gives the participation rate on the same year for the women with no disease. The two curves provide the same information as a naive estimator (i.e., without matching). We see that the probability to work is slightly decreasing before the cancer onset, strongly decreases between one year before and one year after the onset of cancer and regularly decreases afterwards. The probability is almost constant for the women with no disease. There is a clear reduction of the employment probability around the cancer date. However, this plot fails to control for three important elements. First, the effect on employment will obviously depend on the observable heterogeneity of the individuals in the sample (age, education, past health, past employment status). Second, the effect depends on unobservable heterogeneity and we need to use our panel dimension to eliminate it. Third, the effect depends on time since the participation varies with both the business cycle and the progress of medicine. Our methodology addresses all these points.

3. Methodology

Health and labour data can be limited by the data collection process. Administrative data have the advantage of avoiding important drawbacks. Four major problems are often encountered in the applied literature. The first problem relates to the simultaneity between labour and health; the second one relates to the memory and justification biases of the respondents; the third problem is the response rate and the fourth problem is unobserved heterogeneity. We account for the simultaneity

issue both through the time dimension and the non-declarative nature of our data. Consider the time dimension first. We have annual data over the whole career of the people, so that there is no time aggregation, the order of the health and labour events is well known so that no simultaneity should come from aggregation over time. Second, breast cancer is not a standard choice variable like participation in a labour market program, so that the causality is clearer than with the evaluation of labour market policies. Indeed, contrary to more health behavior related cancer (such as lung cancer), we assume that breast cancer is a more random cancer. Outside genetic predisposition, some risk factors exist for breast cancer like obesity, alcohol, tobacco, night work, being pregnant after 30, but at a low risk level. Third, breast cancer is likely to come from factors which are fairly stable over time, like a genetic predisposition, so that they can be accounted for by an individual fixed effect. Fourth, cancer can also come from factors closely related to our observable variables, like age, the education proxy, past health and past labour activity. We account for it with matching.

Consider now the administrative nature of the data. On the one hand we avoid the memory bias since the data is recorded automatically at the time of each event by the administration and, on the other hand, there is no justification bias coming from the respondent because the data is not declarative. Last, the response rate is 100% since the administration records automatically the employment status and the medical information for everyone every year.

We have decided to use the difference in differences (henceforth DiD) method with coarsened exact matching for the following reasons. First, DiD is well known to allow for the elimination of unobserved correlated heterogeneity ("fixed effects"). Second, we use matching in order to weaken the parallel trend assumption associated with DiD (Abadie, 2005). We compare individuals with the same values of all the matching variables so that the trends are assumed parallel inside each of the cells defined by the levels of these variables and not globally like in standard DiD. The spirit of the method is close to Jeon (2017).

We also provide two other estimators in appendix. The first estimator includes the lagged activity dummy among the matching variables. It has the consequence to transform the double differences into simple differences.[7] Therefore this estimator ignores the trend issue and focuses on matching. It should measure what happens when we consider a model with an individual fixed effect and matching. The second estimator is the simple difference in differences estimator. It ignores matching and, therefore, assumes parallel trends for all the individuals in the sample. It should inform us about the importance of this assumption. Overall the first estimator should highlight the contribution of

[7] This is because $y_{i,t_i-1} = y_{j,t_i-1}$.

matching and the second one the consequences of differencing (see Lechner, 2013). We find that matching is very important for our results and that the parallel trend hypothesis does not hold.

There remains to choose the matching method. Here the size of our data set played an important role because a large reservoir of not treated increases the matching possibilities. We have chosen coarsened exact matching rather than propensity score matching (henceforth, PSM) for the following reasons. First, the PSM is often used because, in some data sets, there are not enough observations to perform an exact matching. This can be related to the data set size. Our data is interval or categorical, except for the birth year, and the size of the data set is large enough to allow for an exact matching on the birth year itself. We reach matching rates close to 100% with an exact matching on the birth year. Second, two individuals that have the same matching variable values also have the same PSM because it is a function of these variables: exact matching encompasses PSM. Third, more technical, on panel data, the issue of unobserved heterogeneity in the propensity score is often avoided because there is no simple estimation method which allows for estimating the individual fixed effects (Hsiao, 1996), apart from the conditional Logit model which can involve a large loss of information.[8] The main difficulty lies in the estimation of the fixed effects in the propensity score equation, which are needed to compute a proper prediction of the treatment probability. As a consequence a large part of the literature controls for observable heterogeneity only, by including time-constant variable in the regression and estimates a PSM with no standard panel data fixed effects.[9] Direct matching methods fix the problem by avoiding the estimation of the PSM.

The model. Since we have discrete variables we would like, ideally, to estimate a fully saturated model relating the employment dummy $y_{i,t}$ to the observables $\boldsymbol{X}_{i,t}$ and the treatment $T_{i,t}$ (see Angrist and Pischke, 2009). To this model, we would add a disturbance $u_{i,t}$ including a correlated ("fixed") effect α_i, a time effect $\beta_{0,t}$ and an idiosyncrasic error term $\varepsilon_{i,t}$:

$$y_{i,t} = p_{i,t} + u_{i,t}$$
$$p_{i,t} = f_i(\boldsymbol{X}_i) + \beta_{1,t}(\boldsymbol{X}_i) + \gamma_i(t - t_i) \times T_{i,t}$$

with $u_{i,t} = \alpha_i + \beta_{0,t} + \varepsilon_{i,t}$. The first term $f_i(\boldsymbol{X}_i)$ is a function relating $\boldsymbol{X}_i$ to the employment probability $p_{i,t}$. It includes all the time constant variables (like the education level, the year of birth and the starting wage class) and the time varying variables taken one year before the cancer onset (the

[8] More precisely we would need to drop the observations of the individuals whose cancer status does not change over time, that is all the people without any cancer at all. Since they constitute the natural comparison group, the conditional Logit may not lead to the best propensity score estimate, even in the favourable case where the time dimension would be long enough to allow for the estimates of the fixed effects.

[9] One alternative could be to use the Chamberlain (1980) method which involves some restrictions about unobserved heterogeneity, but no application seems to have been done in this context yet.

past sick leave ratio and the past stable employment ratio). It also includes all the interactions terms between all these variables. It would clearly be very difficult to estimate such a model because of the potential multicollinearity problems created by many cross products. Therefore, we will proceed by matching. Since two individuals who share the same levels of the discrete variables also share the same cross products of these variables (and any function of it), matching is the simplest way to eliminate the outcome differences which come from the individual variables.

The second term $\beta_{1,t}(\boldsymbol{X}_i)$ includes the interactions between the observable discrete variables and the time effect. It is therefore related to the "parallel trend" assumption in the difference in differences literature. We include this interaction term in order to relax the parallel trend assumption. More precisely, we will assume parallel trends for individuals who share exactly the same value of the covariates only. If there are dozens of combinations of the discrete variables, then our model will be equivalent to a model with dozens of time trends. We will eliminate this component by combining differencing with matching.

The third term γ_i is the effect of breast cancer on the outcome variable of individual i. We wish to estimate its average value, denoted γ, the average effect of the treatment on the treated (henceforth, ATT). $T_{i,t}$ is a dummy variable equal to 1 if there is a breast cancer ($\forall t \geq t_i$), 0 otherwise ($\forall t < t_i$). We allow for some forms of heterogeneity in this parameter of interest: it is allowed to vary with the value of the matching variables and the time elapsed since the onset of cancer ($t - t_i$). Indeed, the effect of breast cancer may be different one year after the onset of cancer and five years after. It may also vary with age or the date of cancer onset. The effects according to the matching variables are easily obtained since the estimator can be decomposed according to these variables. Therefore, we will simply report the conditional ATTs defined below. The variation over time of the ATTs will be obtained by taking longer differences, we estimate $\gamma(k)$, $\forall k \geq 1$.

Estimation. In order to identify the impact of health events, we need to account for two types of quantities: on one hand, the difference in histories between those women who experienced cancer and other women and, on the other hand, the variations in the labour history of one woman before and after cancer. The DiD method with a coarsened exact matching allows us to estimate the effect of cancer by controlling both for the observable individual variables and unobservable individual heterogeneity, including when the latter is correlated with the observable individual variables.

The outcome variables are the annual activity dummies corresponding to the employment status (stable and unstable). One can interpret our analysis as an assessment of the impact of breast cancer on these employment status dummies. The estimator is defined by:

$$\hat{\gamma}(k) = \frac{1}{I}\sum_{i \in I}\left(\left(y_{i,t_i+k} - y_{i,t_i-1}\right) - \frac{1}{J(i)}\sum_{j \in J(i)} \left(y_{j,t_i+k} - y_{j,t_i-1}\right) \right)$$

Where I denotes both the treated set of indices and their cardinal, and $J(i)$ denotes both the set of indices of i's twins and their cardinal:

$$J(i) = \left\{ j : \left(t_j > t_i + k\right) \cap \left(\boldsymbol{X}_i = \boldsymbol{X}_j\right) \right\}$$

where $\boldsymbol{X}_i$ is the vector of the matching variables, including the classes defined from the continuous variables. The first condition $t_j > t_i + k$ means that we match the women diagnosed with a breast cancer with women who did not have any cancer or chronic disease before $t_i + k + 1$. When someone does not experience cancer (or another long-term disease), we use the convention $t_j = \{+\infty\}$. This comparison group is made of people with no significant health problem. This can be problematic when we compare our results with the previous literature, since other choices have been made. A current convention takes people with no cancer as a comparison group. In theory, the two conventions should give close results in our application, because we use officially recognized chronic diseases, which are attributed restrictively, especially before the age of 65. In practice, we have made additional estimations with the “no cancer” comparison group. The estimates, reported in appendix (Table A.2) are very similar to the one obtained with our definition.

Table 3: Effect of a breast cancer, DiD with matching

Difference-in-differences with dynamic matching estimates. Matching variables: year of birth (exact matching), first job relative income class (r, 4 levels), past health class (3 levels), past stable employment class (3 levels).

Treated: number of women with a breast cancer.
Matching rate: percentage of treated women that could be matched.
Average number of matches: the average number of twins with who treated women have been matched.
Average in t-1: proportion of women employed one year before the cancer onset.
ATT: Average effect of the treatment on the treated.
ASE: Asymptotic standard errors.
t : Cancer onset date.

Estimation	Treated	Matching rate	Average number of matches	Employment			Stable		Unstable	
				average in t-1	ATT	ASE	ATT	ASE	ATT	ASE
All Sample										
t+1	2662	100%	144	0.798	-0.091*	0.003	-0.118*	0.003	0.027*	0.003
t+2	2389	100%	141	0.797	-0.075*	0.003	-0.079*	0.003	0.004	0.003
t+3	2159	100%	138	0.796	-0.074*	0.003	-0.076*	0.004	0.002	0.003
t+4	1914	100%	135	0.795	-0.076*	0.004	-0.071*	0.004	-0.005†	0.003
t+5	1671	100%	135	0.807	-0.100*	0.004	-0.077*	0.004	-0.023*	0.003
$\text{Age} \leq 48$										
t+1	1238	100%	197	0.897	-0.113*	0.004	-0.158*	0.005	0.045*	0.004
t+2	1116	100%	194	0.900	-0.088*	0.004	-0.108*	0.004	0.019*	0.005
t+3	1034	99.9%	187	0.897	-0.077*	0.004	-0.097*	0.005	0.019*	0.005
t+4	926	100%	181	0.900	-0.075*	0.004	-0.083*	0.005	0.008	0.005
t+5	835	100%	180	0.895	-0.096*	0.005	-0.089*	0.005	-0.007	0.005
$\text{Age} > 48$										
t+1	1424	100%	97	0.711	-0.072*	0.004	-0.083*	0.004	0.011*	0.004
t+2	1273	100%	93	0.707	-0.063*	0.005	-0.053*	0.005	-0.009*	0.004
t+3	1125	100%	92	0.703	-0.071*	0.005	-0.057*	0.005	-0.014*	0.004
t+4	988	100%	90	0.696	-0.076*	0.006	-0.060*	0.006	-0.016*	0.004
t+5	836	100%	89	0.719	-0.104*	0.006	-0.065*	0.007	-0.039*	0.005
$\text{Onset} \leq 1990$										
t+1	147	100%	172	0.857	-0.124*	0.008	-0.137*	0.008	0.012*	0.003
t+2	146	100%	171	0.849	-0.141*	0.006	-0.097*	0.010	-0.043*	0.008
t+3	148	99.3%	172	0.850	-0.106*	0.008	-0.118*	0.012	0.012	0.009
t+4	143	100%	172	0.853	-0.093*	0.008	-0.095*	0.010	0.002	0.007
t+5	145	100%	169	0.855	-0.129*	0.009	-0.122*	0.010	-0.007	0.008
$1991 \leq \text{Onset} \leq 1999$										
t+1	786	100%	140	0.833	-0.098*	0.005	-0.100*	0.005	0.002	0.005
t+2	786	100%	139	0.830	-0.079*	0.005	-0.084*	0.005	0.004	0.005
t+3	787	100%	138	0.825	-0.092*	0.005	-0.092*	0.006	0.000	0.005
t+4	785	100%	136	0.825	-0.098*	0.005	-0.087*	0.006	-0.011*	0.005
t+5	793	100%	136	0.825	-0.107*	0.006	-0.075*	0.007	-0.032*	0.005
$\text{Onset} \geq 2000$										
t+1	1729	100%	142	0.776	-0.085*	0.003	-0.125*	0.004	0.040*	0.004
t+2	1457	100%	137	0.774	-0.065*	0.004	-0.074*	0.004	0.008*	0.004
t+3	1224	100%	132	0.771	-0.058*	0.005	-0.060*	0.005	0.002	0.005
t+4	986	100%	126	0.762	-0.055*	0.005	-0.055*	0.006	0.000	0.005
t+5	733	100%	125	0.778	-0.087*	0.006	-0.070*	0.006	-0.017*	0.005

* significant at the 5% level. † significant at the 10% level.

4. Results

We use the DiD with matching method in order to eliminate the effect of the confounding variables. We present four estimations: full sample, by age class (below and above the median) and by date of cancer onset (before 1990, from 1991 to 1999, from 2000). We use the two following lagged variables: the fraction of years with a stable employment situation ($c \leq 0.5, 0.5 < c \leq 0.7,\ c > 0.7$) and the fraction of years with at least one sick leave quarter ($h = 0, 0 < h < 0.06, h > 0.06$). The matching rates are always close or equal to 100% due to the large reservoir of twins.

Table 3 presents the employment rates one year before the diagnosis of cancer (column “average in t-1”). Considering the whole sample, about 80% of the women were employed before the cancer onset. One year after cancer, the reduction in employment (the ATT) equals 9.1pp (percentage points). After five years, it reaches 10pp. We observe a U-shaped effect due to the very high opportunity cost of the first year related to heavy treatments and potentially, at medium term, the change in work-life-leisure balance preference with advanced age (increase in leisure preference and in constraint between market production and household production).

Some differences appear according to the age at cancer. One year before the cancer onset the younger women (less than 48) work more often (90%) than the older women (71%). The employment rate reduction one year after cancer (the ATT) reaches 11.3pp for the younger women and 7.2pp for the older women, so that the effect is stronger in the short run for the younger women. On the long run however, all women face the same employment rate reduction (-9.6pp for the younger women and -10.4pp for the older women). Cancers at earlier ages are often diagnosed with delay due to the absence of screening programmes and the low incidence of the disease in that age range (less than 5% of breast cancers occur before 40 in 2000’s). Moreover, cancers at earlier ages can receive a worse prognosis due to the higher frequency of triple negative cancers. However, we can assume that the young women may have more ability to return to work earlier than do older women, who also tend to be more discriminated against.

Several alternative explanations might account for the consequences of a patient's age at cancer onset. First, the nature of comorbidities and treatments might differ depending on the age of occurrence. Second, breast cancer occurring at older ages can be especially disabling and prevent women from maintaining a job (undergoing hormonotherapy, which is more frequently applied for post-menopausal women, is given for five years and often entails unpleasant side effects). Third, as more women approach retirement age, the opportunity cost of exiting employment decreases. Fourth, the decline in the probability of employment for older women may be explained by the "double penalty"

phenomenon that can lead to amplified effects of exiting the labour market. Traditional analyses of investments in human and health capital can be enhanced by accounting for changes in preferences or age-related discrimination, which has been particularly noted for older workers (e.g., Datta Gupta and Larsen (2010)).

We find stronger differences according to the cancer onset date. We distinguish three intervals: before 1990, between 1991 and 1999 and from year 2000. A cancer before 1990 reduces the employment rate by 12.9pp after five years; a cancer during the 1990's reduces this figure by 10.7pp and, from 2000, by 8.7pp. There is a gain of about two employment points per decade. Considering the employment types, we find that cancer affects the stable employment status. The differences are statistically significant at the 5% level.

There is a generational effect, for which several explanations could be found. Among them, the improvement in medical treatments for cancer in recent years could 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-year-old women. This trend reversed after 2005. The incidence of invasive cancers grew slowly from 1990 to 1996 and then more sharply beginning in 1996, primarily among women aged 50-74 years, before declining in 2004. Finally, cancers in an advanced stage at diagnosis decreased after a peak in the early 2000s. Changes in incidence likely reflect the combined influence of several factors (screening and diagnostic techniques, and perhaps risk factors). In addition, organized mammography screening for breast cancer was widespread in France in 2004. This programme allows all women aged 50 to 74 to have a free mammogram and clinical breast exam once every two years. The treatments have also improved and it is likely that, like in developed countries, the main effect on the decline of mortality rates is due to the medical advances[10]. Until 1990 the basic treatment was mainly surgery followed by chemotherapy combined with radiotherapy. Since the 2000s, the treatments have improved with better a focus (e. g. conformational radiotherapy, personalized medicine –see e.g. trastuzumab in case of her2 neu positive distant cancers from the early 2000s and in case of her2 neu non distant cancers from 2004-2005 in France) and less functional sequels (sentinel ganglion techniques arriving in France from 2001-2002). Thus, for breast cancer, the 5-year survival increased from 81% in 1990 to 89% in 2002 in France.

The appendix provides 7 other estimators. Table A.1 presents the variants which uses the same control group as in Table 3. First, we take the reference date 2 years before cancer. Our results are not altered. The same conclusion is reached if we use the symmetric lag (t-k,t+k) estimator of Chabé-Ferret (2015). Introducing the lagged endogenous variable among the matching variables only sligthly reduces the

[10] The interpretation of Narod et al. (2015) according to which the decline in mortality due to breast cancer in USA is mainly connected to advanced treatment can also be retained in the case of France.

effect of breast cancer by 1pp (-9.2 after five years instead of -10pp). Finally, the difference-in-differences estimator without matching provides a much stronger effect of breast cancer (-22.9 pp after five years), but it is sensitive to the parallel trends hypothesis and cannot be trusted. Table A.2 reports the estimates obtained with a different control group than in Table 3. The reference group of Table 3 is composed of the people with no breast cancer or chronic illness before the end of the difference (t+k, k=1 to 5). Three other control groups are used: no breast cancer before the end of the difference (disregarding chronic illnesss), no breast cancer at any date, and no breast cancer or chronic illness at any date. Our results are unaltered by these changes in the control groups.

In comparison with other recent studies, using similar methodologies, our results highlight stronger negative effects on the employment situation in the french private sector. A first comparison can be done with the two recent papers from Heinesen and Kolodziejczyck (2013, 2016). In these two papers, the authors use also administrative data and merge them from several registers, including the Danish Cancer Registry which covers all cases of cancer in Denmark since 1943. In a first article (Heinesen and Kolodziejczyck, 2013), the authors focus only on the private sector whereas they focus on both public and private sectors in their second article (Heinesen and Kolodziejczyck, 2016).

The cancer group consists of all women 30-60 in the year of diagnosis with breast cancer in the period 2000–2004 (Heinesen, Kolodziejczyck, 2013) or 2000-2006 (Heinesen and Kolodziejczyck, 2016) who did not have any cancer before this period. The women of their sample had to survive at least until the end of the third year after diagnosis and had to be wage earners two years before. The control group has the same age and no cancer diagnosis at the time of the comparison. They found that cancer reduces the employment probability by 5pp to 10pp after three years, and more strongly for the less educated.

Jeon (2017) uses administrative Canadian data from 1991 to 2002, pooling together women and men. The health variable includes cancers diagnosed for the first time, with about 25% of breast cancer. The main result is that the employment rate decreases by 3 points in t+1 for cancer survivors. The employment rate decreases from 3.5 pp in t+2 to 4.7 pp in t+3. An age effect is also found close to those found by Moran et al. (2011). For young people (28-54) and all cancers they find a loss of 3-5 pp in the employment rate.

Like in Heinesen and Kolodziejczyck and Jeon's work, we also look at a first diagnosis of cancer. But, as contrary to Jeon, we focus on breast cancer whereas Jeon evaluate a mean effect of cancer on employment outcomes[11]. We also study women for whom the diagnosis can date from the late seventies to the early 2000s and on a longer time horizon (five years instead of three in this literature).

[11] The initial sample in Jeon (2017) represents 2597 cancer survivors (677 breast cancer survivors)

The studies of Heinesen and Kolodziejczyck (2013, 2016) are limited to the 2000–2005 diagnosis period and to a 3-year period following the cancer onset. They cannot highlight a generation effect and they do not estimate a long lasting effect. Using a control group including the people with no cancer, Heinesen and Kolodziejczyck (2013) find that the probability of being employed decreases by 4.4 pp in t+1, 5.7 pp in t+2 and 6.7pp in t+3. After changing the control group by including the people diagnosed with a breast cancer 5 years later than treatment group, they obtain similar results. One of the strength of our work is also the size of treated group. Even if we do not benefit from a Cancer Registry as Heinesen and Kolodziejczyck (2013, 2016), we observe 2662 breast cancer survivors one year after the onset cancer and we can follow 1671 women five years after the diagnosis.

5. Conclusion

For the first time in France, we estimate the effect of breast cancer on labour market outcomes in the private sector and up to five years survivors. Compared with the international literature, we analyze a very long diagnosis period, allowing for the analysis of cohort effects. We also estimate the employment outcomes after to five years after, instead of 3 in the other papers.

We use difference in differences with matching techniques. We then use, as a control group, all the women with neither a cancer nor a long term disease at the time of the cancer onset. We provide robustness checks including the use of another control group and other estimation techniques. We find significant effects of breast cancer on labour market outcomes. We find that the onset of breast cancer lead to a loss of 10 pp of employment rate one year after the diagnosis. This detrimental effect remains stable 5 years after the disease. We also point out that premature age, young generations and high initial income appear as protective factors. Our study is original in the french context. First, the estimation of the effect of cancer on professional situations covers a long-term period from one to five years after diagnosis in the french case. In addition, it relies on administrative data to identify the careers of a large sample of private sector employees. Moreover, the sample size permits performing a DiD analysis with coarsened exact matching and defining a rigorous control group that exploits the panel dimension of the data. Finally, we examine a relatively young female population, which reinforces the relevance of the analysis of career paths in this population.

Our study confirms, for the first time in France, with these methods and this time horizon, the detrimental effect of breast cancer on employment for different cohorts of breast cancer survivors. The proportion of individuals who have completed at least one quarter of employment decreases substantially after the onset of cancer (9 pp after one year). This effect is long-lasting because it remains at 10 pp five years after cancer onset.

It is obviously difficult to precisely compare our results with those of other studies because of the differences in data and methods and cross-country differences in labour market structure, public funding of cancer's cost and sick leaves. Bearing this in mind, our findings are rather different from those of Jeon's (2017), who focus on different sites of cancers, and are more similar to those of Moran, Short and Hollenbeak (2011) and of Heinesen and Kolodziejczyk (2013, 2016). Using US data, Moran, Short and Hollenbeak (2011) focus on a young population, as we do, and estimate the effect of surviving cancer on long-term employment outcomes (2-6 years post-diagnosis) in Pennsylvania. However, the sample size of breast cancer survivors is small (230 women), the period of diagnosis is short (1997-1999) and they measure the average effect of cancer, thereby limiting the scope and the comparability of the findings. Using Danish administrative register data, Heinesen and Kolodziejczyk (2013) estimate the effects of breast cancer on labour market outcomes for three-year survivors. They find a smaller effect than we do. It is likely, however, that the generosity of the french health insurance system explains partly the more frequent occurrence of non employment. In comparison with those previous papers, we are also able to control for the whole past career of the people before the diagnosis of the cancer. To some extent, the duration of the past career before the diagnosis can be seen as a specific protective factor against the deleterious effects of breast cancer on employment. Once controlled for it, we find a stronger effect of breast cancer. Nevertheless, the most interesting and innovative findings concern the potential generational effect related to the medical advances. This result would deserve to be explored in greater depth.

Finally, several limitations of our study should be noted. In addition to that related to the definition of cancer (specific to this study), the data do not allow the identification of the cancer stage, the severity and the type of treatment.

In terms of policy implications, a policy goal could attempt to improve job sustainability by reducing the impact of unwanted professional shocks such as breast cancer. A second policy goal should be to increase the (re)integration of cancer survivors by intervening at the beginning of the professional career as recommended by Health Authority (HAS) reading group (HAS 2019). It would avoid a permanent exclusion from the labour market and negative effects of unwanted non-employment or non-employment on health status. In order to fight against the negative effect of an advanced age of cancer onset on employment, it seems efficient to encourage the organized screening (the rate participation of the women aged 50 to 74 is only 50% in France) but also to promote a prevention and a screening visit at age 25 (proposed by the French Health Ministry in 2018).

References

Abadie, A., 2005. Semiparametric Difference-in-Differences estimators. Review of Economic Studies 72(1),1-19.

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics. Princeton University Press, ISBN 978-0-691-12034-8.

Barnay, T., Ben Halima, M. A., Duguet, E., Lanfranchi, J., Le Clainche, C., 2015. La survenue du cancer : effets de court et moyen termes sur les situations professionnelles. Economie et Statistique 475-476, 157-186.

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

Bradley, C. J., Bednarek, H. L., Neumark, D., 2002. Breast cancer survival, work, and earnings. Journal of Health Economics 21(5), 757-779.

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), 833-849.

Bradley, C. J., Neumark, D., Bednarek, H. L., Schenk, M., 2004. Short-term effects of breast cancer on labor market attachment: Results from a longitudinal study. Journal of Health Economics 24, 137-160.

Bradley, C. J., Neumark, D., Luo, Z., Bednarek, H. L., 2007. Employment-contingent health insurance, illness, and labor supply of women: Evidence from married women with breast cancer. Health Economics 16 (7), 719-737.

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), 739-747.

Carlsen, K., Badsberg, J. H., Dalton, S. 2014. Unemployment among breast cancer survivors. Scandinavian Journal of Public Health 42(3), 319-28.

Carlsen K., Oksbjerg, D.S., Frederiksen, K., Diderichsen, F., Johansen, C., 2008. Cancer and the risk for taking early retirement pension: a Danish cohort study. Scandinavian Journal of Public Health 36, 117–125.

Chabé-Ferret S., 2015. Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. Journal of Econometrics, 185(2015) , pp. 110-123.

Chamberlain G., 1980. Analysis of Covariance with Qualitative Data. The Review of Economic Studies, Vol. 47, No. 1, Econometrics Issue (Jan., 1980), pp. 225-238

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

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

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

Damkjaer, L. H., Deltour, I., Suppli, N. P., Christensen, J., Kroman, N. T., Johansen, C., Dalton, S. O., 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), 274-281.

Datta Gupta, N., Larsen, M., 2010. The impact of health on individual retirement plans: Self-reported versus diagnostic measures. Health Economics 19(7), 792-813.

De Boer, A. G., Taskila, T., Ojajarvi, A. van Dijk, F. J., Verbeek, J. H., 2009. Cancer survivors and unemployment: a meta-analysis and meta regression. JAMA 301 (7), 753-762.

Degorre A., Martinelli D., Prost C., 2009. Accès à l'emploi et carrière : le rôle de la formation initiale reste déterminant. *Bilan Formation Emploi*, pp. 9-24.

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), 765-771.

Duguet, E., Le Clainche, C., 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), 49-79.

Eichenbaum-Voline, S., Malavolti, L., Paraponaris, A., Ventelou, B., 2008. Cancer et activité professionnelle. La Revue de l'OFCE 104, 105-134.

Eunmi, A., Cho, J., Shin, D. W., Park, B. W., Ahn, S. H., Noh, D. Y., Nam, S. J., Lee, E. S., Yun, Y. H., 2009. Impact of breast cancer diagnosis and treatment on work-related life and factors affecting them. Breast Cancer Research and Treatment 116 (3), 609-616.

Farley, P., Vasey, J., Moran, J., 2008. Long-term effects of cancer survivorship on the employment of older workers. Health Services Research 43 (1), 193-210.

Feuerstein, M., Todd, B. L., Moskowitz, M. C., Bruns, G. L., Stoler, M. R., Nassif, T., Yu, X., 2010. Work in cancer survivors: A model for practice and research. Journal of Cancer Survivorship, Research and Practice 4, 415-437.

Gregg, P., Tominey, E., 2005. The wage scar from youth unemployment. The Centre for Market and Public Organisation Working Paper Series No. 04/097, Department of Economics, University of Bristol, UK.

Grossman, M., 1972. On the concept of health capital and the demand for health. Journal of Political Economy 80 (2), 223-255.

HAS, 2019. Santé et prévention de la désinsertion professionnelle des travailleurs. Recommandation de bonne pratique, février, 199 pages.

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

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, 247-283.

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), 605–654.

Heinesen, E., Kolodziejczyk, C., 2013. Effects of breast and colorectal cancer on labour market outcomes--average effects and educational gradients. Journal of Health Economics 32, 1028-1042.

Hsiao, C., 1996. Logit and Probit Models. In Matyas, L., Sevestre, P., eds. The Econometrics of Panel Data., Kluwer Academic Publishers, ISBN 0-7923-3787-5, ch. 16, 410-428.

INCA, 2014. Annual report, French Cancer Institute.

INCA, 2015. Annual report, French Cancer Institute.

Jagsi, R., Hawley, S., Abrahamse, P., Li, Y., Janz, N., Griggs, J., Bradley, C.J., 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), 1854-1862.

Jeon, S-H., 2017. The Long-Term Effects of Cancer on Employment and Earnings. Health Economics, 26: 671–684.

Jeon, S-H., Pohl, V., 2017. Health and work in the family: Evidence from spouses' cancer diagnoses. Journal of Health Economics, 52, 1-18

Johnsson, A., Fornander, T., Rutqvist, L. E., Olsson, M., 2011. Work status and life changes in the first year after breast cancer diagnosis. Work 38, 337-346.

Jooste, V., Grosclaude, P., Remontet, L., Launoy, G., Baldi, I., Molinie, F., Arveux, P., Bossard, N., Bouvier, A. M., Colonna, M., 2013. Unbiased estimates of long-term net survival of solid cancers in France. International Journal of Cancer. 132 (10), 2370-2377.

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 diagnosis. Annals of Economics and Statistics 107-108, 239-266.

Kolodziejczyk, C., Heinesen, E., 2016. Labour market participation after breast cancer for employees from the private and public sectors: Educational and sector gradients in the effect of cancer. Economics and Human Biology, 21, 33–55.

Krueger, A.B., Cramer, J., Cho, D., 2014. Are the long-term unemployed on the margins of the labor market? Brookings Papers on Economic Activity, Spring, 229-280.

Lechner, M., 2013. Treatment effects and panel data. Working paper, University of St Gallen, No 2013-14.

Le Rhun B., Monso O., 2015. De l'utilité d'obtenir son diplôme pour s'insérer : l'exemple des brevets de technicien supérieur. Economie et statistique, n° 478-479-480.

Lindbohm, M. L., Kuosma, E., Taskila, T., Hietanen, P., Carlsen, K., Gudbergsson, S., Gunnarsdottir, H., 2011. Cancer as the cause of changes in work situation. Psychooncology., 20, 805-812.

Marino, P., Sagaon, T. L., Malavolti L.., Le Corroler-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, 1-10.

Mehnert, A., 2011. Employment and work-related issues in cancer survivors. Critical Reviews in Oncology Hematology 77, 109-130.

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

Mujahid, M. S., Janz, N. K., Hawley, S. T., Griggs, J. J., Hamilton, A. S., Graff, J., Katz, S. J., 2011. Racial/ethnic differences in job loss for women with breast cancer. Journal of Cancer Survivorship, Research and Practice 5, 102-111.

Narod S. A., Iqbal J., Miller A. B. (2015), Why have breast cancer mortality rates declined ? *Journal of Cancer Policy* 5 : 8-17

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), 322-330.

Observatoire Sociétal des Cancers, 2014. Cancer du Sein : se reconstruire après une mastectomie. Annual report 2014.

Paraponaris, A., Teyssier, L. S., Ventelou, B., 2010. .Job tenure and self-reported workplace discrimination for cancer survivors 2 years after diagnosis: Does employment legislation matter? Health Policy. 98, 144–155.

Petersson L.M., Wennman-Larsen, A., Nilsson, M., Olsson, M., Alexanderson, K., 2011. Work situation and sickness absence in the initial period after breast cancer surgery. Acta Oncologica 50(2), 282-288.

Pourquier, H., 2000. Le traitement des cancers du sein : ses progrès et ses limites. Cancer/Radiothérapie 4 : 248-52.

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), 285-296.

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

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

Yarker, J., Munir, F., Donaldson-Feilder, E., Hicks, B., 2010. Managing Rehabilitation: Competency Framework for Managers to Support Return to Work. BOHRF, London.

Appendix 1 - Additional estimates

Table A.1 – Estimates from alternative methods

ATT: Average effect of the treatment on the treated. ASE: Asymptotic standard errors. t : cancer onset. Total sample estimates.

Estimates	Employment		Stable		Unstable	
	ATT	ASE	ATT	ASE	ATT	ASE
Reference in t-2						
(t-2,t+1)	-0.093*	0.003	-0.137*	0.003	0.044*	0.003
(t-2,t+2)	-0.075*	0.003	-0.097*	0.003	0.022*	0.003
(t-2,t+3)	-0.077*	0.003	-0.093*	0.004	0.016*	0.003
(t-2,t+4)	-0.080*	0.004	-0.088*	0.004	0.008*	0.003
(t-2,t+5)	-0.103*	0.004	-0.095*	0.004	-0.008*	0.003
Symmetric lags						
(t-1,t+1)	-0.098*	0.003	-0.127*	0.003	0.029*	0.003
(t-2,t+2)	-0.075*	0.003	-0.097*	0.003	0.022*	0.003
(t-3,t+3)	-0.073*	0.004	-0.088*	0.004	0.016*	0.003
(t-4,t+4)	-0.085*	0.004	-0.082*	0.004	-0.002	0.004
(t-5,t+5)	-0.103*	0.004	-0.094*	0.005	-0.009*	0.004
Difference-in-differences with matching on the lagged outcome						
(t-1,t+1)	-0.090*	0.002	-0.124*	0.003	0.033*	0.002
(t-1,t+2)	-0.071*	0.002	-0.084*	0.003	0.013*	0.002
(t-1,t+3)	-0.072*	0.003	-0.082*	0.003	0.010*	0.002
(t-1,t+4)	-0.073*	0.003	-0.072*	0.003	-0.001	0.002
(t-1,t+5)	-0.092*	0.003	-0.077*	0.004	-0.015*	0.002
Difference-in-differences without matching						
(t-1,t+1)	-0.127*	0.007	-0.168*	0.009	0.041*	0.008
(t-1,t+2)	-0.136*	0.008	-0.158*	0.009	0.022*	0.008
(t-1,t+3)	-0.157*	0.009	-0.182*	0.010	0.025*	0.009
(t-1,t+4)	-0.182*	0.011	-0.205*	0.012	0.023*	0.009
(t-1,t+5)	-0.229*	0.012	-0.236*	0.013	0.007	0.010

* significant at the 5% level. † significant at the 10% level.

Table A.2 – Estimates with alternative control groups

ATT: Average effect of the treatment on the treated. ASE: Asymptotic standard errors. t : cancer onset. Total sample estimates.

Reference control group used in the other tables: no chronic disease or breast cancer before

Estimates	Employment		Stable		Unstable	
	ATT	ASE	ATT	ASE	ATT	ASE
no breast cancer before**						
(t-1,t+1)	-0.096*	0.003	-0.124*	0.003	0.029*	0.003
(t-1,t+2)	-0.077*	0.003	-0.080*	0.003	0.003	0.003
(t-1,t+3)	-0.075*	0.003	-0.076*	0.004	0.001	0.003
(t-1,t+4)	-0.077*	0.004	-0.070*	0.004	-0.007†	0.004
(t-1,t+5)	-0.100*	0.004	-0.074*	0.004	-0.026*	0.003
no breast cancer at all**						
(t-1,t+1)	-0.096*	0.003	-0.124*	0.003	0.029*	0.003
(t-1,t+2)	-0.077*	0.003	-0.080*	0.003	0.003	0.003
(t-1,t+3)	-0.075*	0.003	-0.076*	0.004	0.001	0.003
(t-1,t+4)	-0.077*	0.004	-0.070*	0.004	-0.007†	0.004
(t-1,t+5)	-0.100*	0.004	-0.074*	0.004	-0.026*	0.003
no chronic disease or breast cancer at all						
(t-1,t+1)	-0.097*	0.003	-0.127*	0.003	0.030*	0.003
(t-1,t+2)	-0.080*	0.003	-0.084*	0.003	0.004	0.003
(t-1,t+3)	-0.079*	0.003	-0.082*	0.004	0.002	0.003
(t-1,t+4)	-0.082*	0.004	-0.076*	0.004	-0.005	0.004
(t-1,t+5)	-0.106*	0.004	-0.081*	0.004	-0.024*	0.003

* significant at the 5% level. † significant at the 10% level. ** the estimates are identical after rounding at 3 figures after the points (there are differences when we round with more figures).

Appendix 2 - Estimation of the standard errors

Imbens and Rubin (2015, chap. 18, p. 425) indicate that exact matching using more than one match is possible when two conditions are fulfilled. First, there should be a large pool of possible control units. Second, increasing the number of matches should not increase the average covariate discrepancy between pairs. Both conditions are fulfilled in our case. First, we can match 2,477 women with a breast cancer with more than 180,000 women without any chronic disease. Second, our matching is exact on all the covariates so that there is no additional discrepancy on the covariates involved by the method.

The computation of the standard errors is simplified by the two following properties of our computation method: we define groups of treated according to the value of their covariates. Since matching is exact, there can be no twin in common between two different groups of treated, so that we can use an independence assumption between groups. Moreover, since the treated all share the same value of the covariates within each group, they have the same twins and their performance is compared to the same twins' average performance inside each group. The ATT estimator, denoted $\hat{\gamma}(k)$, can be written as follows:

$$\hat{\gamma}(k) = \frac{1}{I}\sum_{i \in I}\left(\Delta d_i - \frac{1}{J(i)}\sum_{j \in J(i)} \Delta d_j\right)$$

with Δd_i=$d_{i,t_i+k} - d_{i,t_i-1}$ and Δd_j=$d_{j,t_i+k} - d_{j,t_i-1}$. The previous formula defines the DiD estimator. If two treated individuals have the same matching variables, and if their treatment occurs at the same date, they will be matched with exactly the same twins, and hence the same mean will be subtracted from their outcome variable. We regroup the treated according to their matching variables and treatment date. Let $g \in G$ be a specific vector regrouping the matching variables and the treatment date; the set of all the treated individuals in the matching group g (and their number) is defined by:

$$I(g) = \{i \in I : (\boldsymbol{X}_i, t_i) = g\}, g \in G$$

and we let $J(g)$ denote the common matching twins' set of the treated in group g (and their number). By definition, the $I(g)$ sets define a partition of the treated set $I = \cup_g I(g)$, $I(g) \cap I(g') = \emptyset \; \forall g \neq g'$. Therefore the ATT can be rewritten as follows:

$$\hat{\gamma}(k) = \frac{1}{I}\sum_{g \epsilon G}\sum_{i \in I(g)}\left(\Delta d_i - m_g\right)$$

with $m_g = J(g)^{-1}\sum_{j \in J(g)} \Delta y_j$ the twin's mean within group g. Simplify the sum, we obtain

$$\hat{\gamma}(k) = \frac{1}{I}\sum_{g\epsilon G}\left\{\sum_{i\in I(g)} \Delta d_i - I(g)m_g\right\}$$

In order to compute the variance of this estimator, we make the standard independence assumption between the d_i's. First, the observations in the groups g are independent of one another because they have neither a treated nor a twin in common. We obtain

$$\mathrm{V}\big(\hat{\gamma}(k)\big) = \frac{1}{I^2}\sum_{g\epsilon G}\mathrm{V}\left(\sum_{i\in I(g)} \Delta d_i - I(g)m_g\right)$$

Second, the d_i's are independent of the m_g's because they are computed from different individuals. We obtain

$$\begin{aligned}\mathrm{V}\big(\hat{\gamma}(k)\big) &= \frac{1}{I^2}\sum_{g\epsilon G}\left\{\mathrm{V}\left(\sum_{i\in I(g)} \Delta d_i\right) + I(g)^2\mathrm{V}\big(m_g\big)\right\} \\ &= \sum_{g\in G}\left(\frac{I(g)}{I}\right)^2\left\{\mathrm{V}\big(m_g^T\big) + \mathrm{V}\big(m_g\big)\right\}\end{aligned}$$

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