

L'impact du cancer du sein sur les trajectoires professionnelles

une évaluation sur données administratives

Présentation: E. Duguet¹²³⁵
T. Barnay¹²⁵ C. Le Clainche³⁴

¹Université Paris Est – ERUDITE ²TEPP FR CNRS 3435

³Centre d'Études de l'Emploi et du Travail

⁴Université de Lille 2 – LEM

⁵Institut Santé-Travail Paris-Est

Journée IST-PE, Inégalités de Santé et Travail, 9 Novembre 2017

Sommaire

- 1 Introduction
- 2 Modèle
- 3 Méthode
- 4 Application
- 5 Tableaux

Introduction

Fort accroissement du nombre de cancers

- 109% en France, sur 1980-2012
- âges médians : 67 ans (Femmes) et 68 ans (Hommes) en 2012
- détection plus précoce

Accroissement du nombre d'années requises pour le départ à la retraite

- un plus grand nombre de travailleuses sont encore actives quand le cancer est diagnostiqué
- le cancer interagit plus souvent avec les trajectoires professionnelles
- et l'interaction dure plus longtemps

Introduction

Les cancers

- réduisent la productivité des travailleurs
- les travailleurs réduisent leur offre de travail pendant le temps des soins

Risque accru

- perte d'emploi (inadaptation au poste)
- difficulté à retrouver un emploi après une longue maladie
- moindre investissement en formation permanente

Dans cet article

On étudie

- les effets d'un cancer du sein
- pendant les cinq années qui suivent le diagnostic

sur

- l'emploi (stable, instable)
- le chômage
- les arrêts maladie

Hygie

Sources

- CNAV : données de retraite
- CNAM-TS : données de santé
- nous suivons 225,331 femmes, de leur entrée sur le marché du travail jusqu'en 2008
- 5,161,332 observations

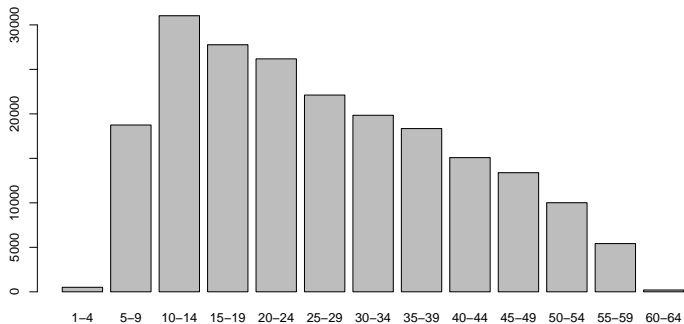
Données disponibles

- individus : sexe, année de naissance, qualification (proxy), emploi, chômage, retraite
- santé : CIM des ALD, arrêts maladie

Nature de l'information

- travail : nombre de trimestre cotisés
- santé : passage en ALD (affection de longue durée)

Nombre d'observations par personne



Données sur le cancer du sein

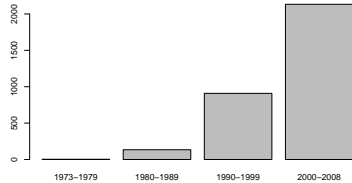
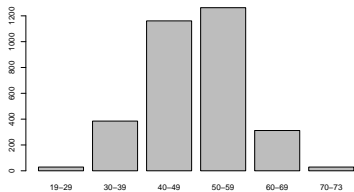
Hygie

- 2,889 femmes avec un cancer du sein
- 19,050 avec une maladie chronique (incluant les autres cancers)
- 203,392 sans maladie chronique

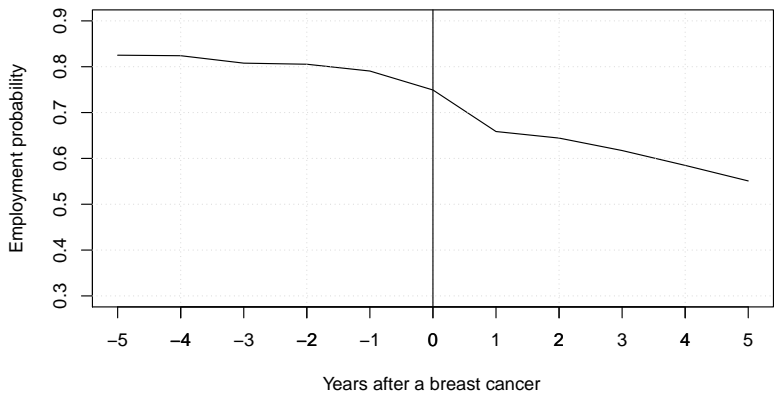
Femmes avec un cancer du sein

- âge médian en ALD : 48 ans
- plus jeune que la population totale (68 ans)
- objectif : étudier des femmes actives

Age au cancer et date de diagnostic



Probabilité d'emploi, estimation naïve avant-après



Modélisation

On étudie la variation de la participation au marché du travail après un cancer du sein (variable indicatrice). Sans cancer, on aurait :

$$y(x, t) = f(x) + \alpha + \beta(t) + \phi(x, t) + \sum_{\tau=t_0+1}^t \gamma_0(\tau - t_0)$$

avec

- y : variable étudiée (indicatrice d'emploi)
- $f(x)$: variables individuelles observables (année de naissance, qualification,...)
- α : hétérogénéité individuelle inobservable
- $\beta(t)$: effets temporels
- $\phi(x, t)$: interactions entre x et t
- $\gamma_0(\tau - t_0)$: évolution de la variable étudié y après $\tau - t_0$ années sur le marché du travail, en l'absence de cancer
- t_0 : date d'entrée sur le marché du travail

Contrefactuel

L'équation précédente

- définit le contrefactuel: ce qui se passe en l'absence de cancer
- on l'estime à partir des données des personnes qui n'ont pas encore eu de cancer ou une autre maladie chronique
- l'équation définit également les composantes à éliminer pour estimer l'effet spécifique du cancer du sein

Dans le cas général :

- $\gamma_0(e)$, $e = t - t_0 \geq 1$ représente le sentier de référence et on modélise le cancer comme une perturbation de ce sentier de référence
- la durée e commence avec la date d'entrée sur le marché du travail t_0

Contrefactuel

Les travailleuses qui n'ont pas encore été en ALD

- sans ALD sur toute la période d'étude
- ou qui auront une ALD plus tard

On effectue un appariement dynamique

- la même personne peut être “traitée” à une date et avoir été “contrôle” à une date antérieure
- en début de période, la même population de référence est utilisée pour toutes les ALD

Sentier d'évolution

Les invididus entrent sur le marché du travail en t_0

- la variable de résultat varie de $\gamma_0(t - t_0)$ entre $t - 1$ et t
- la variation cumulée à la date t vaut :

$$\gamma_0^+(t - t_0) = \sum_{\tau=t_0+1}^t \gamma_0(\tau - t_0)$$

- quand le cancer du sein survient, à la date t_1 , une nouvelle période commence. La **dévi**ation du sentier normal est notée $\gamma_1(t - t_1)$
- le sentier "perturbé" par le cancer à la date t est constitué de deux couches, qui commencent à des dates différentes :

$$\gamma(t) = \gamma_0^+(t - t_0) + \gamma_1^+(t - t_1)$$

Quantities of interest

On utilise la littérature sur l'estimation des effets des traitements :

- on estime les effets du cancer du sein e années après son diagnostic, $\gamma_1(e)$
- on élimine les autres sources de participation au marché du travail:

$$f(x), \alpha, \beta(t), \phi(t, x), \gamma_0^+(t)$$

On utilise la méthode de différence des différences avec appariement:

- différence: $\alpha, \beta(t), \gamma_0^+(t)$
- appariement: $f(x)$
- différence et appariement: $\phi(t, x)$

Sélection des contrôles

On choisit des travailleuses

- sans ALD à la date de comparaison ($t_{1,j} > t_{1,i} + e$)
- avec la même valeur des variables observables ($x_j = x_i$)
- présentes aux mêmes dates : $t_{1,i} - 1$ et $t_{1,i} + e$, $e \geq 1$

Différence des différences

- Pour une traitée $i \in I$, on utilise la performance moyenne de tous les contrôles disponibles $j \in M_1(i)$

$$DD_i(e) = y_i(t_{1,i} + e) - y_i(t_{1,i} - 1) \\ - \frac{1}{M_1(i)} \sum_{j \in M_1(i)} (y_j(t_{1,i} + e) - y_j(t_{1,i} - 1))$$

où $M_1(i)$ est l'ensemble des jumeaux de i (et leur nombre) quand on évalue $\gamma_1^+(e)$.

- L'estimateur est donnée par:

$$DD(e) = \frac{1}{I} \sum_{i \in I} DD_i(e)$$

où I est le nombre de traitées (et leur nombre)

- la variance peut se calculer par une formule exacte

Application

Variables d'évaluation

- variables indicatrices : emploi, chômage, arrêt maladie
- l'information d'origine est en trimestres, donc les indicatrices peuvent être toutes égales à 1 la même année

Variables d'appariement

- année de naissance: assez d'observations pour un appariement exact
- proxy de qualification : ratio salaire de début de carrière/salaires médian de la même année, en 4 classes
- santé passée : nombre d'années avec un trimestre d'arrêt maladie/nombre total d'années (en $t_i - 1$)
- emploi passé : nombre d'années avec un emploi stable / nombre total d'années (en $t_i - 1$)
- emploi stable: une année avec 4 trimestre d'emploi (pas de chômage)
- emploi instable : une année avec moins de 4 trimestres d'emploi, ou avec au moins un trimestre de chômage
- chômage : année avec uniquement des trimestres de chômage

Application

Estimation

- entre $l = 2,662$ ($e = 1$) et $l = 1,671$ ($e = 5$) femmes avec un cancer du sein et une différence $(t_i - 1, t_i + e)$ disponible dans la base
- taux d'appariement: 99.9%-100%
- 135 contrôles en moyenne par traité
- effet négligeables du cancer du sein sur la retraite (âge médian de 48 ans au diagnostic)
- effet négligeable des décès sur l'attrition

Estimation 1/4, Différence des différences avec appariement

Estimation	Traitées	Appar.	# contrôles	Emploi			Emploi stable		Emploi instable	
				Traitées $t_i - 1$	Effet	E.T.	Effet	E.T.	Effet	E.T.
Ensemble										
$t_i + 1$	2662	100%	144	0.798	-0.091*	0.003	-0.118*	0.003	0.027*	0.003
$t_i + 2$	2389	100%	141	0.797	-0.075*	0.003	-0.079*	0.003	0.004	0.003
$t_i + 3$	2159	100%	138	0.796	-0.074*	0.003	-0.076*	0.004	0.002	0.003
$t_i + 4$	1914	100%	135	0.795	-0.076*	0.004	-0.071*	0.004	-0.005†	0.003
$t_i + 5$	1671	100%	135	0.807	-0.100*	0.004	-0.077*	0.004	-0.023*	0.003
Age ≤ 48										
$t_i + 1$	1238	100%	197	0.897	-0.113*	0.004	-0.158*	0.005	0.045*	0.004
$t_i + 2$	1116	100%	194	0.900	-0.088*	0.004	-0.108*	0.004	0.019*	0.005
$t_i + 3$	1034	99.9%	187	0.897	-0.077*	0.004	-0.097*	0.005	0.019*	0.005
$t_i + 4$	926	100%	181	0.900	-0.075*	0.004	-0.083*	0.005	0.008	0.005
$t_i + 5$	835	100%	180	0.895	-0.096*	0.005	-0.089*	0.005	-0.007	0.005
Age > 48										
$t_i + 1$	1424	100%	97	0.711	-0.072*	0.004	-0.083*	0.004	0.011*	0.004
$t_i + 2$	1273	100%	93	0.707	-0.063*	0.005	-0.053*	0.005	-0.009*	0.004
$t_i + 3$	1125	100%	92	0.703	-0.071*	0.005	-0.057*	0.005	-0.014*	0.004
$t_i + 4$	988	100%	90	0.696	-0.076*	0.006	-0.060*	0.006	-0.016*	0.004
$t_i + 5$	836	100%	89	0.719	-0.104*	0.006	-0.065*	0.007	-0.039*	0.005

Effet: Effet moyen du traitement sur les traitées. E.T.: Ecart-type. *: significatif à 5%. †: significatif à 10%.

Estimation 2/4, Différence des différences avec appariement

Estimation	Traitées	Appar.	# contrôles	Emploi			Emploi stable		Emploi instable	
				Traitées $t_i - 1$	Effet	E.T.	Effet	E.T.	Effet	E.T.
$r \leq Q_1$										
$t_i + 1$	648	100%	168	0.730	-0.141*	0.006	-0.161*	0.007	0.020*	0.007
$t_i + 2$	582	100%	166	0.725	-0.120*	0.007	-0.114*	0.007	-0.006	0.007
$t_i + 3$	534	100%	164	0.727	-0.133*	0.007	-0.124*	0.008	-0.008	0.007
$t_i + 4$	479	100%	164	0.722	-0.132*	0.008	-0.108*	0.009	-0.024*	0.008
$t_i + 5$	426	100%	167	0.739	-0.172*	0.008	-0.138*	0.009	-0.035*	0.007
$Q_1 < r \leq \text{Me}$										
$t_i + 1$	517	100%	137	0.793	-0.128*	0.004	-0.117*	0.006	-0.011 [†]	0.006
$t_i + 2$	466	100%	129	0.794	-0.097*	0.005	-0.075*	0.006	-0.023*	0.007
$t_i + 3$	424	100%	123	0.788	-0.093*	0.007	-0.066*	0.008	-0.028*	0.008
$t_i + 4$	375	100%	121	0.789	-0.083*	0.007	-0.054*	0.009	-0.029*	0.008
$t_i + 5$	327	100%	122	0.792	-0.088*	0.008	-0.041*	0.007	-0.047*	0.009
$\text{Me} < r \leq Q_3$										
$t_i + 1$	655	100%	130	0.855	-0.088*	0.006	-0.148*	0.007	0.060*	0.006
$t_i + 2$	582	100%	127	0.854	-0.071*	0.006	-0.080*	0.007	0.009 [†]	0.005
$t_i + 3$	510	100%	128	0.857	-0.070*	0.006	-0.072*	0.008	0.002	0.006
$t_i + 4$	437	100%	121	0.851	-0.067*	0.007	-0.075*	0.008	0.008	0.005
$t_i + 5$	370	100%	118	0.859	-0.092*	0.009	-0.082*	0.009	-0.010*	0.005
$r > Q_3$										
$t_i + 1$	842	100%	143	0.808	-0.032*	0.004	-0.063*	0.005	0.030*	0.004
$t_i + 2$	759	100%	140	0.810	-0.029*	0.005	-0.053*	0.006	0.024*	0.005
$t_i + 3$	691	100%	134	0.810	-0.019*	0.006	-0.047*	0.006	0.029*	0.005
$t_i + 4$	623	100%	131	0.814	-0.034*	0.006	-0.050*	0.007	0.016*	0.005
$t_i + 5$	548	100%	130	0.832	-0.057*	0.007	-0.049*	0.008	-0.008	0.005

Effet: Effet moyen du traitement sur les traitées. E.T.: Ecart-type. *: significatif à 5%. †: significatif à 10%.

Estimation 3/4, Différence des différences avec appariement

Estimation	Traitées	Appar.	# contrôles	Emploi			Emploi stable		Emploi instable	
				Traitées $t_i - 1$	Effet	E.T.	Effet	E.T.	Effet	E.T.
Naissance \leq 1952										
$t_i + 1$	1517	100%	107	0.716	-0.067*	0.003	-0.081*	0.004	0.014*	0.003
$t_i + 2$	1424	100%	105	0.724	-0.066*	0.004	-0.063*	0.004	-0.003	0.004
$t_i + 3$	1326	100%	105	0.729	-0.072*	0.004	-0.073*	0.005	0.002	0.004
$t_i + 4$	1227	100%	105	0.733	-0.081*	0.005	-0.073*	0.005	-0.007†	0.004
$t_i + 5$	1107	100%	106	0.761	-0.106*	0.005	-0.074*	0.005	-0.032*	0.004
Naissance $>$ 1952										
$t_i + 1$	1145	100%	194	0.906	-0.123*	0.004	-0.167*	0.005	0.044*	0.005
$t_i + 2$	965	100%	194	0.905	-0.088*	0.004	-0.102*	0.005	0.014*	0.005
$t_i + 3$	833	100%	190	0.904	-0.077*	0.005	-0.080*	0.006	0.003	0.006
$t_i + 4$	687	100%	189	0.905	-0.067*	0.005	-0.067*	0.006	0.000	0.006
$t_i + 5$	564	100%	192	0.897	-0.088*	0.007	-0.084*	0.007	-0.005	0.006

Effet: Effet moyen du traitement sur les traitées. E.T.: Ecart-type. *: significatif à 5%. †: significatif à 10%.

Estimation 4/4, Différence des différences avec appariement

Estimation	Traitées	Appar.	# contrôles	Emploi			Emploi stable		Emploi instable	
				Traitées $t_i - 1$	Effet	E.T.	Effet	E.T.	Effet	E.T.
Diag. ≤ 1990										
$t_i + 1$	147	100%	172	0.857	-0.124*	0.008	-0.137*	0.008	0.012*	0.003
$t_i + 2$	146	100%	171	0.849	-0.141*	0.006	-0.097*	0.010	-0.043*	0.008
$t_i + 3$	148	99.3%	172	0.850	-0.106*	0.008	-0.118*	0.012	0.012	0.009
$t_i + 4$	143	100%	172	0.853	-0.093*	0.008	-0.095*	0.010	0.002	0.007
$t_i + 5$	145	100%	169	0.855	-0.129*	0.009	-0.122*	0.010	-0.007	0.008
1991 \leq Diag. \leq 1999										
$t_i + 1$	786	100%	140	0.833	-0.098*	0.005	-0.100*	0.005	0.002	0.005
$t_i + 2$	786	100%	139	0.830	-0.079*	0.005	-0.084*	0.005	0.004	0.005
$t_i + 3$	787	100%	138	0.825	-0.092*	0.005	-0.092*	0.006	0.000	0.005
$t_i + 4$	785	100%	136	0.825	-0.098*	0.005	-0.087*	0.006	-0.011*	0.005
$t_i + 5$	793	100%	136	0.825	-0.107*	0.006	-0.075*	0.007	-0.032*	0.005
Diag. ≥ 2000										
$t_i + 1$	1729	100%	142	0.776	-0.085*	0.003	-0.125*	0.004	0.040*	0.004
$t_i + 2$	1457	100%	137	0.774	-0.065*	0.004	-0.074*	0.004	0.008*	0.004
$t_i + 3$	1224	100%	132	0.771	-0.058*	0.005	-0.060*	0.005	0.002	0.005
$t_i + 4$	986	100%	126	0.762	-0.055*	0.005	-0.055*	0.006	0.000	0.005
$t_i + 5$	733	100%	125	0.778	-0.087*	0.006	-0.070*	0.006	-0.017*	0.005

Effet: Effet moyen du traitement sur les traitées. E.T.: Ecart-type. *: significatif à 5%. †: significatif à 10%.