The impact of health events on individual labor market histories : the message from difference in differences with exact matching

Emmanuel Duguet, Christine Le Clainche

The impact of health events on individual labor market histories : the message from difference in differences with exact matching

Emmanuel DUGUET*
Université Paris Est, ERUDITE and TEPP
and
Christine LE CLAINCHE†
Centre d'Études de l'Emploi
LAMETA UMR 5474 ENS Cachan

January 2012

Abstract

We studied the effect of health events (accidents and chronic diseases) on the occupation probabilities at the individual level, while accounting for both correlated individual and time effects. Using difference-in-differences with exact matching estimators, we found that health events have a strong impact on individual labor market histories. The workers affected by a healt event have a stronger probability of entering inactivity and a lower probability of keeping their jobs. We also found that the less qualified workers, women, and workers with short term jobs are the most negatively affected by health events.

JEL : C14, I10, J20, J60.

*ERUDITE, Université Paris Est, Mail des Mèches, 61 avenue du Général de Gaulle, 94010 Créteil Cedex. Email : emmanuel.duguet@u-pec.fr.

†Adress : Centre d'Etudes de l'Emploi, "Le Descartes I", 29 promenade Michel Simon, 93166 Noisy le Grand Cedex. Email : christine.leclainche@cee-recherche.fr.

1 Introduction[1]

This article presents a study of the impact of health events (an accident or a chronic disease) on participation in the labor market. This topic usually raises causality issues, since health affects labor, and labor affects health. In order to adress this problem, we use a data set that provides annual information about labor and the occurrence of health events. This allows us to distinguish clearly the period before and after the health events, and to perform a difference in differences analysis. The time dimension is especially important in disentangling the joint causality between health and labor.

This topic had already been studied with reference to seniors in France, where there is now a significant literature about the "health shocks" that provides evidence on early exits out of the labor market through early retirement (Barnay (2005), Blanchet, Debrand (2007), Barnay, Debrand, (2007), Debrand (2007), Behaghel, Debrand, Roger (2011) for France, for example). However, there is no comparable evidence for younger populations. In this paper, we exclude the retired workers and focus on the impact of health events on the labor market history of young and middle-aged workers.

One possible outcome of this analysis is the possibility of reducing the economic and social costs implied by accidents and chronic diseases. There has been no economic evaluation in France of the costs implied by poor health, mainly because there are no aggregated accounts associated with prevention policies as a whole (Heijink, Noethen, Renaud, Koopmanshap, Polder (2008), Cour des Comptes (2011)). So, a better knowledge of the effects of health events on employment history could allow us to better target health measures in relation to work, by including health issues within the employment histories of workers.

We use the survey " Health and Professional Histories" (Santé et Itinéaires Professionnels, SIP), collected between 2006 and 2007. It is a representative sample of individuals from 19 to 74 years old, which describes both health events and employment history, on a yearly basis. This is the first time that such information is available in France, and it allows us to perform a dynamic analysis of health and occupation. In order to estimate the

[1]This research has benefited financial support from Drees and Dares (Research Department of the French Ministry of Labour and Social Affairs). We thank for their comments on earlier versions of this paper: Chantal Cases, Thierry Debrand, François Gardes, Véronique Simonnet , Nicolas Sirven, Christophe Starzec and Claude Mathieu. We also thank the participants to the following seminars and conferences : Health and Safety at Work Conference (Universita Cattolica Milano, July 2010); Journée de Microéconomie Appliquée (University of Angers, June 2010), Journée des Economistes de la Santé Français (University of Lyon, December 2010), Journées d'Etudes sur l'Hétérogénéité des Comportements Individuels (ANR-06-Blan-0140, University of Paris I, April 2011), Journée Santé-Travail, (Chaire Santé, Risque Assurance, University of Paris-Dauphine, September 2011) and the Third French Econometrics Conference (Aix-Marseille University, Aix en Provence, December 2011).

impact of accidents and the chronic diseases on occupation, we implement a method of difference-in-differences with exact matching. This method is non parametric and allows us to identify the causality running from health events to occupation. We find that health events have significant and important effects on participation in the labor market.

The paper is organized as follows : in the second section, we summarize the evidence from the economic literature. The third section presents the data used in this application and is followed by a presentation of the econometric method in the fourth section. The results are presented in the fifth section, after which we reach the conclusion.

2 Review of the literature

2.1 Effects of health events in the labor market

Poor health has negative and significant economic and social impacts in the labor market, especially due to the exits out of the labor market it induces. For example, Chaupain-Guillot and Guillot (2010) evaluate, from Eurostat data, the direct and indirect costs induced by inactivity periods due to ill health at 90 billion euros in Europe, that is about 1% of the GDP of European countries. Since the 1990s, there has been a growing literature dealing with the impact of health on labor market participation, earnings and wages, following the seminal theoretical model of health capital by Grossman (1972) and its Mincerian extensions. In such models, a higher stock of health capital is expected to increase earnings and wages because it allows workers to increase the number of hours worked and also because a higher health capital involves higher productivity. Most of the empirical studies based on these theoretical grounds use cross-section data which do not take into account unobservable heterogeneity. An overview of the empirical literature by Currie and Madrian (1999) shows that poor health reduces the ability to work and has significant effects on wages, labor force participation and job choice but the results are sensitive to the measurement of health chosen, and within the surveyed literature relatively few studies are based on longitudinal surveys.

The dynamic analysis of the impact of health on labor market histories is just beginning. Using six waves of the British Household Panel Survey (BHSP, 1991-1996), Contoyannis and Nice (2001), produced one of the first studies shedding light on the effects of overall and on psychological health on wages, by taking into account unobservable heterogeneity. They used a single fixed effects wage equation with self-reported health indicators among the regressors. They found that reduced psychological health has an effect on hourly wage for men, while excellent self-reported health for women has a significant impact on their hourly wage. Another study by Cai and Kalb (2006) includes some dynamic characteristics concerning the labor market, using Australian data from the Hilda data base (2001). They use a self-reported health measure, based on the limitations of everyday activities, and a measure based on the SF 36, which is a measure of general health and well-being, producing scores on eight dimensions of health. The endogenous nature of the health status in the labor force participation equation is addressed by estimating the health equation and the labor participation equation simultaneously. The results show that better health increases the probability of labor force participation for all age groups and both genders, but the biggest effect is obtained for the older groups and for women. For the French case, most of the studies concentrate on seniors except those carried out by Tessier and Wolff (2005). The latter is conducted from the "Timetable" survey " , collected by the National Institute for Statistics and Economic Studies (INSEE) between February 1998 and February 1999, on a selected sample of 7023 individuals from 25 to 55 years old. Tessier and Wolff (2005) estimate the impact of health on participation in the labor market. They use two measures of health: self-reported health and a measure indicating the existence of chronic disease which the authors consider as "more objective". The simultaneous estimation of a participation equation and of a health equation takes into account the simultaneity between the two variables. The estimations bring to light two main results: the simultaneous estimation of the equations shows that good health has a positive impact on labor market participation and that participation in the labor market has no significant effect on health. The second result is that health significantly affects participation in the labor market from the first years of potential activity.

However, the study by Tessier and Wolff does not focus on the impact of health events on employment trajectories, due to the lack of adequate data. The SIP survey that we use in this study fills this gap.[2] Moreover, with the SIP survey, we also are able to account for individual variables that have been proved to have an important influence on health, such as the childhood living conditions.[3]

2.2 The impact of injuries on work and employment

While the impact of accidents on employment is rather well studied, especially in relation to working conditions[4], the impact of domestic, sport or road accidents on employment has been studied much less. Thesealso deserve to be studied, since the economic and human costs of such accidents

[2] Concerning the links between health and unemployment see Haan, Myck (2009); Böckerman, Ilmakunnas (2009).

[3] For the link with market labor performance, see Lindeboom, Llena, Nozal, van der Klauw (2006). Precarious conditions during childhood have important consequences for health in the future and for labor market participation and unemployment.

[4] See Karasek, Theorell, (1990) ; Reville and Schoeni (2001) ; Wichert (2002); and for France, see Askenazy (2002, 2009) ; Hamon-Cholet, Sandret (2007).

are high and the effects on the trajectories of employment are potentially significant. For example, in 2000, road accidents entailed injuries for 1.3 million individuals in Europe (Moller Dano, 2005). In France, the direct and indirect total costs estimated by the French Office for Road Security (ONISR) are estimated as being equal to 1.3 % of the GDP for 2008.

One of the rare studies dedicated to the microeconomic effects of road accidents on employment shows that, for Denmark, the effects of road accidents are serious for both employment and earnings: the rates of employment are respectively 10 and 8 points lower for injured men and women. Besides, earnings are significantly reduced for men whatever their age, but this applies only to the oldest women (Moller Dano (2005)). In order to obtain these results, the author had to correct the effects of selection associated with the risk of accidents, so as to identify the causal impact of car accidents on earnings and rates of employment. It is indeed necessary to correct the selection bias insofar as young men are reputed to have more more risky behavior regarding cars. Moreover, their earnings are lower than those of older men. Another rare study was conducted by Crichton, Stillman, Hyslop (2011) for New Zealand. With a similar methodology, the authors show that there are strong and negative impacts of accidents (including workplace accidents) on employment and earnings. The authors also found that accidents giving the right to long-term earning compensations due to inability to work have a stronger negative impact on women, older workers and people with low incomes.

The SIP survey that we use includes information on accidents and allows us to deal with the impact of accidents on labor market histories.

3 The Data

The survey "Health and Professional Histories" (SIP) was carried out from November, 2006 until January, 2007, on people from 19 to 74 years old. This survey allows us to identify all the stages of a professional history and to observe the health events occurring over the same period. The survey includes questions about childhood and activity periods. In this paper, we focus on the people between 19 and 59 years old. Since we focus on the relatively young people, we have also discarded all the people with a retirement or pre-retirement period.[5]

In order to avoid simultaneity issues, we account for the following points : first, we include correlated individual and time effects in the model, in order prevent unobserved heterogeneity influencing both health events and labor market participation. Secondly, we focus on the health events that are not related to work. Thirdly, we account for the date at which the health events

[5]We also let the disable people apart since they will be the topic of a future separate study.

occur, and examine their effect on the subsequent labor market history.

Our accident variable will be a yearly dummy variable that indicates whether the individual has had an accident. We exclude two types of accidents : accidents in the workplace and accidents during transport to the workplace. Therefore we keep mostly domestic accidents, sport accidents and those car accidents not related to work.

Our data on chronic disease is based on the health care administration classification ("Sécurité Sociale", in France). Chronic diseases initially self-declared, but the declarations must pass through the definition of long-term diseases provided by the "Sécurité Sociale". It is so because, in France, such diseases benefit from a full reimbursement, so the Sécurité Sociale monitors them carefully. In order to identify the chronic diseases we report on epidemiologists' views of diseases causing limitations (see WHO, IDC) and on the French administrative classification of severe diseases (the so-called "Affections de Longue Durée" or ALD classification). In the SIP survey, the data set is very detailed about the type of disease from which people suffer, in a declarative sense. We have retained : chronic cardio-vascular diseases, cancers, incurable deafness, chronic hearing impairment (tinnitus), severe or chronic lung diseases, severe or chronic liver diseases, severe or chronic rheumatism, diabetes, severe or chronic eye disorders (impossible to correct) ; severe or chronic psychiatric disorders, epilepsy, addictions, AIDS or other severe disease. For these data, we keep a yearly dummy variable that indicates whether the individual has a chronic disease.

This allows us to construct four data sets : first, the accident data set, that includes the individuals that had an accident and no chronic disease ($N = 336$); secondly, the chronic disease data set, that includes the individuals that had a chronic disease and no accident ($N = 800$) . Then come two data sets, that we call the reference data sets, that represent the reservoirs from which we select the individuals that will be used for comparisons with the two previous groups. The third data set includes the individuals that had neither an accident nor a chronic disease and that have the same individual variables as the people in the accident data set ($N = 3228$). The fourth data set includes the individuals that had neither an accident nor a chronic disease and that have the same individual variables as in the chronic disease data set ($N = 3618$) . We present the matching method before comparing these groups further.

4 Methodology

We wish to explore the impact of health events on the employment history of workers. We account for three problems. First, it is likely that health events 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 health event, 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 health events for each individual. Secondly, the individuals in the data set have different ages so that the observation window differs from one individual to another. Therefore we need to match individuals by age, so that we compare workers that had a health event (or not) at the same age and during the same year. Thirdly, we account for correlated individual and time effects since we have panel data. Fourthly, we perform a non parametric estimation, so that no specific functional assumption is made for the health-activity relationship.

4.1 Observable heterogeneity, unobservable heterogeneity and exact matching on observables

The SIP survey provides a detailled dynamic account of two main variables : health events and occupational status. In this section we analyze how it is possible to evaluate the impact of health events (accident or a first chronic disease) on the remaining labor market history. In order to identify successfully the impact of health events we need to account for two types of quantity : on the one hand, the difference in histories between the individuals that experienced a health event and the other individuals; on the other hand, the history variation of one individual before and after the health event. In this section, we show that the difference-in-differences method allows us to estimate the effect of health events 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 left-hand variables of interest are the annual activity dummies corresponding to the four following occupations : inactivity, unemployment (more than 1 year), short-term employment (less than 5 years) and long-term employment (more than 5 years). The means of these dummy variables for one individual give the proportions of time spent in these occupations. One can interpret our analysis as an assessment of the impact of health events on these four occupation probabilities.

Formally, we estimate a model in wich our occupational dummies d_{it} can be broken dowm into four parts :

– Observable individual heterogeneity that represents the right-hand variables available in the survey, denoted X_i. Typically, in the literature, this type of heterogeneity is often represented by a linear econometric model. In this paper, we perform an exact matching on the observable variables so that we do not need to impose any specific functional form relating the left-hand variables to the right-hand variables. Our approach is non parametric. The purpose of the matching is to get rid of the effects of the observable variables on the occupation probabilities. This part of the model is denoted

$f(X_i)$,where f is unknown.

– Unobservable individual heterogeneity that finds its source either in time-constant missing individual variables or corrrelated individual effects. This part of the model is similar to the individual correlated effects in the panel data literature. We cannot exclude the possibility that this heterogeneity is correlated both with the observable individual variables and the occurrence of health events. Difference-in-differences will allow us to get rid of this type of heterogeneity. This component is denoted α_i

– We also add an individual-constant time heterogeneity, possibly correlated to the explanative variables, for the following two reasons. First, we are dealing with periods of several decades, so that the job opportunities clearly differ over the period under study. Secondly, it is also clear that the probability of getting a given disease or of having a specific type of accident can vary over long time periods because of medical or behavioral preventive action. Thirdly, the progress of medicine over long periods also affects the ability to get back to work. Our method allows us to discard these effects, since matching individuals *over the same time period* allows us to difference it out. This component is denoted β_t

– Lastly, we add a standard white noise error term which correspond to the idiosyncratic error in the panel data models, denoted ε_{it}. These errors are uncorrelated with the rest of the model and with each other, and have a mean that can be allowed to vary over time $\mathrm{E}(\varepsilon_{it}) = \mu_t$. We could find the standard panel data model by seting $\mathrm{E}(\varepsilon_{it}) = 0$, but this is not required by our method.

4.2 Econometric model

Our panel is not balanced so that we observe one individual i over a individual specific time period denoted $\left[t_i^-, t_i^+\right]$. Therefore, a health event can occur at any date $t_i \in \left[t_i^-, t_i^+\right]$. In order to evaluate the effect of this health event, we will compare the occupation probabilities over the period $\left[t_i^-, t_i - 1\right]$ and $\left[t_i, t_i^+\right]$. Notice that the dates are integer numbers since we only have annual data. Considering first an individual j with no health event over the period $\left[t_j^-, t_j^+\right]$, we have :

$$d_{jt} = f(X_j) + \alpha_j + \beta_t + \varepsilon_{jt},\ \forall t \in \left[t_j^-, t_j^+\right],$$

where d_{jt} is an occupation dummy. The different occupation dummies can therefore have different $(f, \alpha, \beta, \varepsilon)$. If we consider one individual i that experienced a health event at date t_i, we must distinguish two periods, before and after the health event :

$$d_{it} = f(X_i) + \gamma_i \times \delta(t \geq t_i) + \alpha_i + \beta_t + \varepsilon_{it},$$

with

$$\delta\left(t \geq t_i\right)=\begin{cases}1 & \text{if } t \geq t_i \\ 0 & \text{if } t \leq t_i-1\end{cases}$$

where γ_i is the individual effect of the health event on individual i. It is our parameter of interest. Overal, we can write that :

$$d_{it}=\begin{cases} f\left(X_i\right)+\alpha_i+\beta_t+\varepsilon_{it} & \text{if } t_i^{-} \leq t \leq t_i-1 \\ f\left(X_i\right)+\gamma_i+\alpha_i+\beta_t+\varepsilon_{it} & \text{if } t_i \leq t \leq t_i^{+}\end{cases}$$

In order to estimate an average value for γ_i we must proceed in several steps.

- Step 1 : we compute the occupation differences between the individuals that experienced a health event and the others. The average performance of an individual with a health event *before* the event occurs is equal to :

$$\begin{aligned}\overline{p}_i^0 &= \frac{1}{t_i-t_i^{-}}\sum_{t=t_i^{-}}^{t_i-1} d_{it} \\ &= f\left(X_i\right)+\alpha_i+\frac{1}{t_i-t_i^{-}}\sum_{t=t_i^{-}}^{t_i-1}\left(\beta_t+\varepsilon_{it}\right)\end{aligned}$$

and the average performance of the individuals j that never experienced a health event over the *same* period is equal to :

$$\begin{aligned}\overline{p}_j^0 &= \frac{1}{t_i-t_i^{-}}\sum_{t=t_i^{-}}^{t_i-1} d_{jt} \\ &= f\left(X_j\right)+\alpha_j+\frac{1}{t_i-t_i^{-}}\sum_{t=t_i^{-}}^{t_i-1}\left(\beta_t+\varepsilon_{jt}\right)\end{aligned}$$

where $\left[t_i^{-}, t_i^{+}\right] \subset \left[t_j^{-}, t_j^{+}\right]$ guarantees that we can match the individuals on the same years, so that we can restrict the comparison to $t_j=t_i$ and $t_j^{-}=t_i^{-}$. This condition is equivalent to matching the individuals on the time period before the accident. Further, in order to eliminate the effect of the observable variables, we match our individuals on the right-hand variables so that we define :

$$j \in J\left(i\right)=\left\{j : X_j=X_i\right\}$$

therefore $j \in J\left(i\right) \Rightarrow f\left(X_i\right)=f\left(X_j\right)$ and we can write the occupation difference over the matched individuals :

$$D_{0i}=\overline{p}_i^0-\overline{p}_j^0=\alpha_i-\alpha_j+\frac{1}{t_i-t_i^{-}}\sum_{t=t_i^{-}}^{t_i-1}\left(\varepsilon_{it}-\varepsilon_{jt}\right)$$

unobservable individual heterogeneity clearly drives this difference, so that we need to go further. Notice however that time heterogeneity has been differenced out.

- Step 2 : we compute the same type of difference on the period after the health event. The average performance for an individual with a health event is :

$$\begin{aligned}\overline{p}_i^1 &= \frac{1}{t_i^+ - t_i + 1}\sum_{t=t_i}^{t_i^+} d_{it} \\ &= f\left(X_i\right) + \alpha_i + \gamma_i + \frac{1}{t_i^+ - t_i + 1}\sum_{t=t_i}^{t_i^+}\left(\beta_t + \varepsilon_{it}\right)\end{aligned}$$

and the average performance for a matched individual without a health event over the same period is :

$$\begin{aligned}\overline{p}_j^1 &= \frac{1}{t_i^+ - t_i + 1}\sum_{t=t_i}^{t_i^+} d_{jt} \\ &= \underbrace{f\left(X_i\right)}_{f(X_j)} + \alpha_j + \frac{1}{t_i^+ - t_i + 1}\sum_{t=t_i}^{t_i^+}\left(\beta_t + \varepsilon_{jt}\right),\ j \in J\left(i\right)\end{aligned}$$

so that we get the difference :

$$D_{1i} = \overline{p}_i^1 - \overline{p}_j^1 = \alpha_i - \alpha_j + \gamma_i + \frac{1}{t_i - t_i^-}\sum_{t=t_i^-}^{t_i - 1}\left(\varepsilon_{it} - \varepsilon_{jt}\right),$$

where time heterogeneity has been differenced out.

- Step 3 : we compute the difference of the differences

This gives :

$$\begin{aligned}D_i &= D_{1i} - D_{0i} \\ &= \gamma_i + \frac{1}{t_i^+ - t_i + 1}\sum_{t=t_i}^{t_i^+}\left(\varepsilon_{it} - \varepsilon_{jt}\right) - \frac{1}{t_i - t_i^-}\sum_{t=t_i^-}^{t_i - 1}\left(\varepsilon_{it} - \varepsilon_{jt}\right)\end{aligned}$$

Now, using the assumption $\mathrm{E}\left(\varepsilon_{it}\right) = \mu_t$, we get :

$$\begin{aligned}\mathrm{E}\left(D_i\right) &= \gamma_i + \frac{1}{t_i^+ - t_i + 1}\sum_{t=t_i}^{t_i^+}\left(\mu_t - \mu_t\right) - \frac{1}{t_i - t_i^-}\sum_{t=t_i^-}^{t_i - 1}\left(\mu_t - \mu_t\right) \\ &= \gamma_i,\end{aligned}$$

where both time and individual heterogeneity have been differenced out. The quantity γ_i is the effect of the health event for the individual i on the occupational dummy d_{it}. Our purpose is to estimate its average value over all individuals (or over subsets of individuals) :

$$\overline{\gamma} = \frac{1}{N_h} \sum_{i=1}^{N_h} \gamma_i,$$

where N_h is the number of individuals that experienced a health event.

So far, our expressions involve theoretic expressions so that we need to find empirical counterparts to proceed to estimation. This is the topic of the next section.

4.3 Estimation

We estimate D_i by using its empirical counterpart. Since the relationships are linear, it is an unbiased estimator. However, in order to improve on the efficiency of the estimator, we will not take only one individual j to be matched with i but *all* the individuals j in the data set that can be matched with i. Therefore the probabilities $\overline{p}_j^k,\ k \in \{0, 1\}$ will be estimated by :

$$\begin{aligned} \widehat{p}_{J(i)}^k &= \frac{1}{N_i} \sum_{j \in J(i)} \widehat{p}_j^k \\ \text{with } \widehat{p}_j^0 &= \frac{1}{t_i - t_i^-} \sum_{t=t_i^-}^{t_i - 1} d_{jt} \ \text{ and } \ \widehat{p}_j^1 = \frac{1}{t_i^+ - t_i + 1} \sum_{t=t_i}^{t_i^+} d_{jt} \end{aligned}$$

where

- $N_i = \# J(i)$ is the number of individuals in $J(i)$.
- $\widehat{p}_j^k$ is the empirical occupational probability of one individual j, matched with individual i since $j \in J(i)$
- $\widehat{p}_{J(i)}^k$ represents the average empirical occupational probability of all the individuals that can be matched with individual i. Its variance is lower than the variance of $\widehat{p}_j^k$, and this is why we prefer this estimator to $\widehat{p}_j^k$.

For one individual i, the differences are estimated by :

$$\widehat{D}_{0i} = \widehat{p}_i^0 - \widehat{p}_{J(i)}^0 \text{ and } \widehat{D}_{1i} = \widehat{p}_i^1 - \widehat{p}_{J(i)}^1,$$

therefore we take the mean of these differences over all the individuals that experienced a health event (number N_h):

$$\widehat{D}_0 = \frac{1}{N_h} \sum_{i=1}^{N_h} \widehat{D}_{0i} \ \text{ and } \ \widehat{D}_1 = \frac{1}{N_h} \sum_{i=1}^{N_h} \widehat{D}_{1i}$$

and we finally estimate the average effect of a health event by :

$$\widehat{\gamma} = \widehat{D}_1 - \widehat{D}_0,$$

this is an unbiased estimator for :

$$\mathrm{E}(\widehat{\gamma}) = \overline{\gamma},$$

which is the average theoretical effect of health over all the individuals that had a health event.

It remains to compute the variance of $\widehat{\gamma}$. One difficulty appears. We can match the same individual j (without a health event) with different i individuals (with a health event). Indeed, the only important point is that the observations periods verify $\left[t_i^-, t_i^+\right] \subset \left[t_j^-, t_j^+\right]$ and that the observable variables are the same $(X_i = X_j)$, since the time period of any individual without any health event can be split at will. We use as many individuals in the reference sample as possible. This implies that the empirical quantities within the means are correlated.

One can easily solve this problem by using the bootstrap, since the empirical means are asymptotically pivotal functions (see Horowitz, 1999). We take a large number of draws in order to get as close as possible to the ideal bootstrap estimate : 10000 draws. More precisely we performed the non parametric bootstrap method on the data base that includes both individuals with and without a health event. This implies that our bootstrap estimates also account for the variation of the number of individuals that can be matched. The whole estimation process is therefore replicated after each draw in the original data base. We performed the estimation under SAS.

5 Results

5.1 The matching variables

We matched the individuals on 8 variables : gender (2 classes), age (4 classes), highest degree (3 classes), nationality of the mother (2 classes), nationality of the father (2 classes), country of birth (2 classes), health problems of the parents (2 classes) and separation from the parents during childhood (2 classes). This makes 384 possible combinations. To these combination, we must add matching over the time periods. Therefore, the exact matching method cannot be applied to all the workers. Table 1 presents both the matching percentages and the average individual characteristics of the workers that could be matched or not. We find that 73% of the accident sample can be matched, and 68% for the chronic disease sample. Therefore, we should look at the characteristics of the people that could not be matched. We find that, overall, the individuals that could be matched have almost the same characteristics as the other people, so that the matching impossibilities comes from the differences in the time periods. This is good news, since the individuals that are not in the estimates share comparable individuals variables with the individuals that are in the estimates, so that the applicability of our conclusion should extend beyond the matched sample. A comparison between the accident and the chronic disease samples is also meaningful since it reveals differences between the two heath events. We find that accidents involve younger people (27% under 36 years old, versus 19%) and more men (60% versus 38%) than chronic diseases.

The difference in differences estimations are presented in Table 2 for the accidents and in Table 3 for chronic diseases. It is important to notice two points. First, the sum of the differences always equals zero since the occupation dummies sum to 1. What we measure is therefore similar to a deformation of the occupation probabilities after a health event. Secondly, it is possible to compute these differences for sub-samples of the original data set, provided that there are enough observations. We find that this is the case for gender and education level. Therefore we will examine whether men and women face comparable consequences on the labor market after a health event, and whether a higher education level allows individuals to compensate for the anticipated negative effect of health events on activity.

5.2 Effect of an accident on occupation

Overall, an accident increases the inactivity probability by 5.3%. Compared to the benchmark probability at 10.7%, it means that the workers that had an accident have, on average, an inactivity probability of 10.7%+5.3%=16% instead of 10.7%. This effect is balanced by a comparable negative effect on employment (-6%). If we look in more detail, we find that the whole effect on employment comes from short-term jobs. Therefore, the workers that had long-term jobs are not significantly affected by an accident, while the workers with a short-time job move to inactivity. However, this global result conceals significant differences among genders and education levels.

Considering gender effects, we find that women are more strongly affected than men by an accident. Their inactivity probability increases by 7.2%, instead of 3.8% for men. We also find that the effect of an accident decreases with the education level. The increase in the inactivity probability is 9.2% for primary education, 5.1% for secondary education and 3.6% for tertiary education. This is balanced by a negative effect for employment.

Therefore the effect of an accident is to move workers out of the labor market, mostly women and those in the less qualified jobs.

5.3 Effect of a chronic disease on occupation

Overall, a chronic disease has a strong impact on occupation : we find that the probability of becoming inactive increases by 7.7%. Compared to the benchmark probability of 14.1%, the workers that face a chronic disease see their inactivity probability rising to 21.8%, and their probability of being employed is reduced by 6.7 points. This effect on employment is equally split between short-term jobs (-3.4%) and long-term jobs (-3.2%). However, this overall effect conceals important composition effects concerning gender and education levels.

The lowest education level clearly concentrates the most negative effects on occupation : the inactivity probability increases by 17.2% and the employment probability decreases by a comparable figure. The types of job that are the most affected by a chronic disease are the long-term jobs (-14%) and, in a much lower extent, the short term jobs (-4%). We notice that the number of transition to unemployment is not affected by the appearance of a chronic disease; therefore it is likely that the less educated individuals make a direct transition from long-term jobs to inactivity. In addition to this, their inactivity probability was already high (benchmark : 24.8%) so that the inactivity probability rises to 42% in the case of a chronic disease. Notice that, with our definition, inactivity does not include retirement, so that the previous result is not driven by the older workers that would retire after a chronic disease. The medium education level is associated with a smaller effect of a chronic disease. The inactivity probability increases by 6.8%, almost the half of the lowest education level. This increase corresponds to a reduction of both the probabilities of unemployment and employment. Within employment, only short-term jobs are affected. Therefore, the incidence of a chronic disease does not modify the probability of keeping a long term job, but makes workers move from unemployment and short-term jobs to inactivity. The highest education level is associated with the lowest impact of chronic diseases : the probability of inactivity increases by 3.1% and the other differences are not significant. Overall, the negative impact of chronic diseases on employment is strongly decreasing with the education level.

6 Conclusion

We performed a dynamic analysis of health events on occupation, that accounts for correlated individual and time effects. We find that health events have a significant and negative effect on activity. Health events always reduce the probability of employment and increase the probability of inactivity. We also find that women are more negatively affected than men, and that the negative effects of health events are decreasing with the education level. Comparing accidents and chronic disease, we find that accidents have an important but smaller effect than chronic diseases on inactivity.

Overall, we find that health events tend to deteriorate the situation of the workers that are already the least favored in the labor market. The less qualified workers and women have the strongest increase in inactivity and, for all workers, the most affected are the ones with a short-term job. Since we are studying the first health event, and since we have excluded work-related health events, it should be a causal effect. One consequence is that health events would drive a high number of workers toward the minimum assistance revenues. This calls for policies oriented towards adapting the workplace to the most common health events or towards training the workers so that they can find another job more compatible with their health problems.

References

Askenazy P. (2002), *La croissance moderne*, Economica.

Barnay T. (2005), « Santé déclarée et cessation d'activité », *Revue Française d'Economie*, vol. XX, n° 2, 73-106.

Blanchet D., Debrand T. (2007), "Souhaiter prendre sa retraite le plus tôt possible : Santé, satisfaction au travail et facteurs monétaires", *Economie et Statistique*, n°403-404, 39-62.

Böckerman, P. & Ilmakunnas, P. (2009). "Unemployment and Self-assessed Health: Evidence from Panel Data". *Health Economics*, vol. 18(2), 161-179.

Cour des Comptes (2011), *La prévention sanitaire*, Report, October.

Debrand T., Sirven N. (2009), "What are the Motivations of Pathways to Retirement in Europe: Individual, Familial, Professional Situation of Social Protection Systems", Working Paper, Irdes n° 28. 2009/10.

Behaghel L., Blanchet D., Debrand T., Roger M. (2011), "Disability and Social Security Reforms: The French Case", NBER Working Paper n°17055, May .

Cai L., Kalb G. (2006), "Health Status and Labour Force Participation: Evidence from Australia", *Health Economics*, vol. 15, 241-261.

Case A., Fertig A., Paxson C. (2005), "The Lasting Impact of Childhood Health and Circumstance", *Journal of Health Economics*, vol. 24, 365-389.

Chaupin-Guillot S., Guillot O. (2010), « Les déterminants individuals des absences au travail : une comparaison européenne », Working Paper Beta, n°17.

Crichton S., Stillman S., Hyslop D.(2011), "Returning to Work from Injury: Longitudinal Evidence on Employment and Earnings", *Industrial and Labor Relations Review,* vol. 64 n°4, July.

Contoyannis P., Nice R. (2001), "The Impact of Health on Wages : Evidence from the British Household Panel Survey", *Empirical Economics*, vol. 26, 599-622.

Currie J., Madrian B. (1999), “Health, Health Insurance and the Labor Market”, in Ashenfelter O., Card D.(eds), *Handbook of Labor Economics*, vol. 3, chap. 50, 3309-3416.

Grossman M. (1972), « On the Concept of Health Capital and the Demand for Health », *Journal of Political Economy*, vol. 80, n° 2, pp. 223-255.

Haan P., Myck M. (2009), "Dynamics of Poor Health and Non-Employment", IZA Discussion Paper n° 4154, April.

Hamon-Cholet S, Sandret N. (2007), “Accidents et conditions de travail”, *Premières Synthèses. Premières Informations* n°31.

Heijink R., Noethen M., Renaud T., Koopmanschap M., Polder J., (2008), “ Cost of illness: An international comparison Australia, Canada, France, Germany and The Netherlands”, *Health Policy*, vol 88, n° 1 2008/10, 49-61.

Idler E. L., Benyamini Y.(1997), “Self-rated health and mortality: a Review of twenty – seven communities studies”, *Journal of Health and Social Behaviour*, vol. 38, 21-37.

Karasek R., Theorell T. (1990), *Healthy Work: Stress, Productivity and the Reconstruction of the Working Life*, Basic Books.

Lindeboom M., Llena-Nozal A., van der Klauw B. (2006), “Disability and Work”, Tinbergen Institute Discussion Paper 039/03.

Moller-Dano A.M. (2005), “Road Injuries and Long Run Effects on Income and Employment”, *Health Economics*, vol. 14, n°9, pp. 955-970.

Reville R. T. Schoeni R.F. (2001), “Disabilities from Injuries at Work: the Effects on Earnings and Employment”, Rand Labor and Population Program Working Paper Series, 01-08.

Tessier P., Wolff F. C. (2005), Offre de travail et santé en France, *Economie et Prévision,* n°168 : 17-41.

Trannoy A., Tubeuf S., Jusot F., Devaux M. (2010), “Inequality of Opportunities in health in France: A first Pass”, *Health Economics*, vol. 19 (8), 921-938.

Wadsworth M., Butterworth S. (2006), “Early Life”, in Marmot M., Wilkinson R. (2006), *Social Determinants of Health*, Oxford University Press, Chap. 3, 31-53

Wichert I. (2002), « Job Insecurity and Job Intensification: the Effects on Health and Well-Being"in Burchell B., Ladipo D., Wilkinson F. (eds), *Job Insecurity and Work Intensification*, Routledge

Table 1: Average characteristics of the samples (proportions)

The characteristics of the reference base, regrouping individuals without any health event, are not indicated in the table since we performed exact matching, which implies that the reference data base shares the same characteristics as the matched accident or chronic disease bases. The figures indicate the proportions of individuals for each level of the matching variables. The similarity in proportions between the matched and the not matched individuals indicates that the matching impossibilities came mostly from the differences in time periods between the individuals.

	Accident base		Chronic disease base	
Individual matching variables	**Matched**	**Not matched**	**Matched**	**Not matched**
Percentage	**73%**	**27%**	**68%**	**32%**
Age :				
19-27	0,068	0,076	0,031	0,123
28-36	0,205	0,228	0,161	0,151
37-45	0,318	0,319	0,306	0,274
46-55	0,408	0,377	0,501	0,452
Gender :				
Women	0,375	0,267	0,601	0,630
Men	0,625	0,733	0,399	0,370
Education level :				
Primary	0,190	0,199	0,234	0,151
Secondary	0,438	0,512	0,409	0,397
Tertiary	0,372	0,289	0,358	0,452
Childhood :				
Foreign mother	0,119	0,140	0,150	0,219
Foreign father	0,110	0,127	0,158	0,192
Born in France	0,884	0,892	0,853	0,863
Parents had serious health problems	0,167	0,147	0,191	0,164
Separated from their family	0,149	0,167	0,153	0,123

Table 2: Average effect of an accident on the occupation probabilities

Difference-in-differences with exact matching. The standard errors and the confidence intervals are computed by the bootstrap method with 10000 replications. The confidence intervals need not be symmetric. ** : significant at 5%; * : significant at 10%.

Sample	Occupation	Benchmark probability $\bar{p}_i^0$	Accident effect $\hat{\gamma}$	Standard error	Student	Confidence interval 95%		Confidence interval 90%	
Total	Inactivity	0,107	0,053**	0,014	3,73	0,025	0,081	0,029	0,076
	Unemployment	0,023	0,007	0,010	0,67	-0,012	0,026	-0,010	0,023
	Employment	0,870	-0,060**	0,016	3,66	-0,092	-0,027	-0,086	-0,032
	Incl. less than 5 years	0,334	-0,056**	0,019	2,87	-0,094	-0,018	-0,088	-0,024
	5 years and more	0,536	-0,004	0,023	0,16	-0,048	0,040	-0,040	0,033
Gender: Women	Inactivity	0,202	0,072**	0,027	2,62	0,018	0,126	0,027	0,118
	Unemployment	0,026	-0,007	0,014	0,52	-0,036	0,021	-0,031	0,016
	Employment	0,772	-0,065**	0,029	2,25	-0,121	-0,009	-0,112	-0,018
	Less than 5 years	0,294	-0,045	0,030	1,51	-0,104	0,013	-0,094	0,003
	More than 5 years	0,478	-0,020	0,034	0,58	-0,086	0,046	-0,075	0,036
Gender: Men	Inactivity	0,038	0,038**	0,014	2,75	0,011	0,065	0,015	0,061
	Unemployment	0,018	0,017	0,013	1,29	-0,009	0,042	-0,005	0,038
	Employment	0,944	-0,055**	0,018	3,04	-0,091	-0,019	-0,085	-0,025
	Incl. less than 5 years	0,227	-0,064**	0,026	2,40	-0,115	-0,011	-0,107	-0,020
	5 years and more	0,718	0,008	0,031	0,26	-0,054	0,069	-0,043	0,059
Education: Primary	Inactivity	0,269*	0,092*	0,048	1,93	-0,002	0,183	0,014	0,169
	Unemployment	0,002	0,044	0,027	1,61	-0,010	0,092	-0,003	0,086
	Employment	0,729**	-0,136**	0,050	2,70	-0,233	-0,037	-0,217	-0,053
	Incl. less than 5 years	0,296	-0,057	0,047	1,22	-0,148	0,034	-0,134	0,020
	5 years and more	0,433	-0,079	0,064	1,23	-0,203	0,047	-0,183	0,028
Education: Secondary	Inactivity	0,064	0,051**	0,018	2,80	0,016	0,086	0,021	0,081
	Unemployment	0,040	-0,004	0,018	0,23	-0,040	0,030	-0,033	0,025
	Employment	0,896	-0,047*	0,025	1,88	-0,096	0,002	-0,088	-0,005
	Incl. less than 5 years	0,345	-0,089**	0,029	3,06	-0,146	-0,032	-0,136	-0,041
	5 years and more	0,551	0,042	0,034	1,23	-0,025	0,110	-0,014	0,098
Education: Tertiary	Inactivity	0,079	0,036*	0,021	1,71	-0,005	0,078	0,001	0,071
	Unemployment	0,014	0,001	0,011	0,07	-0,021	0,021	-0,017	0,018
	Employment	0,907	-0,037*	0,021	1,78	-0,078	0,004	-0,071	-0,003
	Incl. less than 5 years	0,340	-0,020	0,032	0,61	-0,084	0,045	-0,073	0,034
	5 years and more	0,567	-0,017	0,034	0,51	-0,085	0,049	-0,075	0,038

Table 3: Average effect of a chronic disease on the occupation probabilities

Difference-in-differences with exact matching. The standard errors and the confidence intervals are computed by the bootstrap method with 10000 replications. The confidence intervals need not be symmetric. ** : significant at 5%; * : significant at 10%.

Sample	Occupation	Benchmark probability $\bar{p}_i^0$	Chronic disease effect $\hat{\gamma}$	Standard error	Student	Confidence interval 95%		Confidence interval 90%	
Total	Inactivity	0,141	0,077**	0,011	7,00	0,056	0,099	0,059	0,096
	Unemployment	0,024	-0,011	0,007	1,62	-0,024	0,002	-0,022	0,000
	Employment	0,835	-0,067**	0,013	5,21	-0,092	-0,042	-0,087	-0,046
	Incl. less than 5 years	0,236	-0,034**	0,013	2,70	-0,059	-0,009	-0,055	-0,013
	5 years and more	0,599	-0,032**	0,015	2,12	-0,062	-0,002	-0,057	-0,007
Gender: Women	Inactivity	0,209	0,088**	0,016	5,55	0,057	0,119	0,062	0,115
	Unemployment	0,027	-0,013	0,009	1,50	-0,031	0,004	-0,028	0,001
	Employment	0,764	-0,075**	0,018	4,24	-0,109	-0,040	-0,104	-0,046
	Less than 5 years	0,243	-0,033**	0,017	1,98	-0,066	-0,001	-0,061	-0,006
	More than 5 years	0,521	-0,042**	0,021	2,02	-0,083	-0,002	-0,076	-0,008
Gender: Men	Inactivity	0,038	0,061**	0,014	4,42	0,034	0,088	0,038	0,083
	Unemployment	0,018	-0,007	0,010	0,69	-0,028	0,013	-0,025	0,010
	Employment	0,944	-0,054**	0,017	3,25	-0,086	-0,021	-0,081	-0,027
	Incl. less than 5 years	0,227	-0,035*	0,019	1,85	-0,072	0,003	-0,066	-0,004
	5 years and more	0,717	-0,019	0,021	0,88	-0,061	0,023	-0,054	0,016
Education: Primary	Inactivity	0,248	0,172**	0,033	5,22	0,107	0,237	0,118	0,227
	Unemployment	0,023	0,013	0,014	0,94	-0,015	0,038	-0,011	0,035
	Employment	0,729	-0,185**	0,034	5,38	-0,253	-0,118	-0,242	-0,129
	Incl. less than 5 years	0,201	-0,044*	0,025	1,75	-0,093	0,006	-0,085	-0,002
	5 years and more	0,528	-0,141**	0,040	3,56	-0,219	-0,062	-0,207	-0,075
Education: Secondary	Inactivity	0,123	0,069**	0,017	3,93	0,034	0,102	0,039	0,097
	Unemployment	0,026	-0,025**	0,012	2,02	-0,049	-0,001	-0,045	-0,005
	Employment	0,851	-0,044**	0,021	2,10	-0,084	-0,003	-0,078	-0,009
	Incl. less than 5 years	0,260	-0,045**	0,022	2,03	-0,090	-0,001	-0,082	-0,008
	5 years and more	0,591	0,001	0,024	0,06	-0,046	0,049	-0,038	0,042
Education: Tertiary	Inactivity	0,096	0,031**	0,012	2,65	0,008	0,054	0,012	0,050
	Unemployment	0,021	-0,011	0,009	1,13	-0,028	0,008	-0,026	0,005
	Employment	0,883	-0,020	0,015	1,37	-0,049	0,008	-0,045	0,004
	Incl. less than 5 years	0,232	-0,016	0,018	0,87	-0,052	0,020	-0,046	0,015
	5 years and more	0,651	-0,004	0,020	0,22	-0,043	0,035	-0,038	0,029