Prolonged (sickness) Absence and Job-Duration: Evidence from Severe Health Shocks

Irene Simonetti
Ca' Foscari University of Venice

10/07/2020

Abstract: Based on new administrative data the paper investigates the role played by the length of sick leave - under severe health conditions - on the subsequent risk of leaving the job. Sickness absence is costly. Although it is an instrument designed by policymakers to prevent the potential income losses related to bad health, it is also a channel through which the employer receive negative signals in terms of productivity. When studying the link between health deterioration and labour market activity many institutional factors come into play. Hence, the Italian institutional setting - traditionally characterised by high levels of employment protection legislation (EPL), especially for open-ended contracts - is a good framework for this type of analysis. Results point out a significant, and negative, relationship between sick leave duration and the likelihood of a job-interruption: an additional week at home increases the risk of exit of about 1.6% each time. Surprisingly, age-related differences and firm dimension do not change the effect of extended sick leave. Overall, our findings question the effectiveness of employment-support and protection measures, especially when chronic conditions strongly discourage individuals' working activity.

Keywords: sick leave, job duration, survival analysis, administrative data
JEL codes: C41, I10, J64

Data Acknowledgment: This paper uses an anonymised release of WHIP&Health data, to be solely used in the research project "Study on the impact of diseases on workers' career" (record 119099 of 20/12/2017), and made available to the author by the Department of Epidemiology, ASL TO3 Grugliasco (Turin, Italy)

1 introduction

Increasing the economic activity of unhealthy workers is one of the most important challenges faced by policy makers. When studying the link between health deterioration and labour market activity many institutional factors come into play. In Italy, a country characterised by high levels of employment protection legislation[1] (EPL) (see OECD, 1999; Boeri and Jimeno, 2005), several national and European directives are in place to facilitate individuals with reduced working capabilities to retain their job: unfair dismissals[2] and workplace adjustments such as reallocation of tasks or changes of contractual working hours are among the possibilities. However, despite their availability, the effective application of these measures depends strongly on the employer's trade-off between firing costs and the expenditure to keep unhealthy people working.

With this in mind, the role played by an 'extended sick leave' and its relationship with the subsequent risk of closing a specific labour contract - especially if it is characterised by a high level of protection - is not straightforward. When people experience severe health shocks such as cardiovascular diseases (CVDs), the time needed to recover and return-to-work can be pretty long. At the same time, sickness absence is often considered a proxy of an individual's level of productivity and thus, a negative signal for employers. Depending on the country-specific sickness insurance regulation, notable costs for both employers and employees may arise. Beyond the direct losses people face when the replacement rate of sickness benefits is lower than the 100% of previous earnings, indirect effects can also occur: the reduction of future expected earnings deriving from lay-off or lack of promotions is often strongly related to the worker's history regarding sick leave (Hesselius (2007), Markussen (2012)). From the employers' perspective instead, the longer the time spent at home by the unhealthy worker, the higher the costs faced. Indeed, together with the sizeable productivity losses and additional expenses derived from sick workers' temporary replacement, employers usually pay - along with the Italian Social Security Institute (INPS, in Italian) - a percentage of earnings during illness periods (the so-called "sick pay").

This paper analyses the role of sick leave *duration*, under specific health conditions, in speeding up the exit from the specific job. On the one hand, the strictness of the Italian employment protection legislation and thus, employer's duties in terms of workplace adjustments, can lead to a positive - or even null - relationship between the two components. On the other hand, a variety of factors such as task difficulties, dismissal, reduced careers opportunities, and lack of employer accommodations would instead explain a shortening of the time left on that job. Although unable to disentangle the multiple channels through which such an early exit may occur, the resulting individuals' and social costs can be very high: discontinuous working careers, an increase of social

[1] According to the OECD Employment Protection Index, on a scale of 5 points (stringent EPL), the *Strictness of employment protection*(regular contracts) along the first decade of 2000, the value for Italy is equal to 2.76; while in the UK is stable to 1.26.

[2] A dismissal is unfair unless it is for a just cause (no notice required) or a justified motive (notice required) (see art. 1 and 3 Act 604/1966 and art. 2119 CC.)

security supports' recipients or early-exit from the labour market, are among the possibilities.

Evaluation of the link between sickness absence and a variety of labour market outcomes has often been impeded by one main challenge: the endogeneity of sick leave arising from many unknown factors, especially underlying health conditions. That is why the largest part of this literature has often focussed on assessing the role of hidden opportunistic behaviours rather than its actual effects on workers' careers. National reforms aiming to create a more efficient system of sickness benefits have often been exploited for this purpose (Johansson et al. (2002), Puhan et al. (2010), Ziebarth (2013)). By sampling workers hit by a severe form of CVD shock, a plausible relationship between the length of absence and people's underlying health conditions comes out. Moreover, retrospective information, covering almost fifteen years before the occurrence of the selected health event, have been extensively explored in order to partially take into account the prior health status as well as individual's labour market attachment. To the best of our knowledge, this is among the first papers exploring the survival rates in a specific labour contract and its relationship with a tricky instrument such as sick leave. Rather than looking at the probability of unemployment as an outcome variable, we claim that the "time until a job-interruption occurs" is better able to capture the effectiveness of the employment protection legislation, especially when open-ended contracts are considered. Finally, besides the need to ensure homogeneity among categories of workers receiving sickness benefits from the Italian Social Security Institute (INPS), the choice of blue-collar workers has also a notable policy interest. According to the European Labour Force Survey data (EU-LFS), the percentage of blue-collar men reporting "own illness or disability" as the main reason of leaving the last job is almost always double than that of white-collars (Figure A2 in Appendix): this is potentially related to the type of jobs, often demanding, carried out by the former group of people.

The research takes advantage of a newly available dataset, called WHIP&Health, that links the work histories of a 7% random Italian population from 1990 to 2012, together with individuals' hospitalisations sourced from the hospital discharge registers. In particular, a set of 1354 male and permanent blue-collar workers has been selected. In addition to being affected by a severe CVD shock in a year between 2003 and 2005, the sample selection also guarantees that they have not experienced similar health shocks in either of the two previous years, reassuring that the time spent at home cannot be the result of severe past illnesses (at least in the recent past). The precise information regarding the time of the CVD shock, the weeks in sick leave and the ending date of the labour contract (if it occurs by the end of the observational window) allows computing the distance between the time an individual return to work and the expiring date. Upon this framework, Cox's proportional hazard models are particularly suitable to describe the direction and the magnitude of the studied relationship. As previously stated, a variety of possible situations may occur, and thus, the direction is not straightforward. The first set of results, where the effect of sick leave is kept constant among different age groups, shows that a unitary increase in the number of weeks

at home is associated with the significant growth of the instantaneous risk of exit from that job of about 1.6%. This finding suggests that, despite the legislative attempts to guarantee protection to unhealthy workers, the Italian labour market does not sufficiently help people with reduced working capabilities to continue working. The second part of the analysis takes advantage of the interactions between sick leave and different age groups to assess, if and how the role of a prolonged absence differs by age. Surprisingly, notable differences do not appear between the three younger age groups and the oldest one, i.e. those between 56 and 64 years old. Irrespectively of age, the experience of a CVD shock has undoubtedly negative consequences on individuals' labour market participation, raising questions about the appropriateness of the Italian EPL. Finally, although the effect of sick leave is found to be similar across firm size, this variable is instead relevant *per se* in shrinking the remaining duration of the labour contract. An easier reallocation of tasks together with different legislations in terms of dismissal costs, are some of the possible explanations.

The following section provides a brief overview of the past literature, while Section 1.3 defines the institutional background upon which the research idea is based. Section 1.4 extensively describes the dataset, the sample selection criteria, the available variables and some related issues. Section 1.5 illustrates the econometric approach and the baseline setup. The empirical results and some sensitivity checks are available in Section 1.6 and 1.7. Finally, the conclusions are in Section 1.8.

2 Literature Review

The relationship between sickness absence measured under severe health conditions, and the time until job-interruption occurs is a rather new field of research. Besides the huge literature exploring the moral hazard behaviours arising from different sickness insurance systems, a variety of additional fields are covered. Some studies focus on its link with the increase of unemployment risk/loss of earnings; some others instead, look at the strength to which employers' accommodations can help disabled people to keep on working as well as their fruitful (or not) implementation. To enrich the subsequent discussion, throughout this section the main findings of all these fields of research will be briefly covered.

In general, many direct and indirect costs can arise from sick leave absence. The formers depend on the replacement rates ensured by the country-specific legislation: the higher this rate is, the lower the income loss faced by ill workers. The indirect costs instead, arise from the losses in future expected earnings both in terms of increased layoff probability or missed careers opportunities. As shown in Schön (2015) by using the German Socioeconomic Panel (GSOEP), sick days are a strong predictor of unemployment. This is also confirmed for Italian workers by Scoppa et al. (2014). The idea is that more time spent at home increases the probability of being tagged as a less productive

worker or a shirker (Hesselius (2007)). Thus, people who are less "absence-prone" are also more likely to remain employed in recession times. According to Markussen (2012), although sick leave should help unhealthy people to recover and go back to their job, it may often be a trap. Interestingly, by exploiting the number of certificates granted for sickness - proxy of the leniency of worker's physician - as an instrument for sick leave, he found that a one percent increase in sick leave is associated with a 1.2 percent reduction in earnings two years later, together with an increase of about 0.5 percentage points in the probability of being employed.

Complementary to these works, there is a wide literature concerning worker absenteeism and its moral hazard peculiarities, in relation to individual and labour characteristics. With respect to gender differences, Barmby et al. (2002) develop an interesting international comparison by using EU-LFS data, showing how women have higher absence rates than men in most of the countries; moreover, as expected, similarities are found among older individuals, who face increasing difficulties also due to the 'ageing process'. On a different perspective, Pfeifer (2013) explores the absenteeism phenomenon among private sector, public sector and self-employed workers. First of all, in line with the research exploiting the pro-cyclical trends of sick leave with the country-specific economic situation (Arai et al. (2005), Askilden et al. (2005), Schön (2015)), he confirms how regional unemployment rates are negatively correlated with the number of absent working days; and it is true for private, public and self-employed workers. Public sector employees, typically characterised by stronger job-protection rules, have the highest rates of absenteeism; they are followed by private-sector workers and finally by self-employed workers. Focusing on Italian public employees, De Paola et al. (2014) found that an Italian Law passed in June 2008, aiming to reduce sick leave compensation and increasing the monitoring of absences, has negatively affected workers' opportunistic behaviours, especially among those facing the highest-earning losses. Similarly, Puhani et al. (2010) evaluate the effects of a reduction in sick pay from 100% to 80% of the wage in Germany; while Johansson et al. (2002) explore Swedish blue-collar workers. Finally, heterogeneities in the type of contract have also been extensively studied by the literature (Leombruni (2011)): workers with temporary contracts take fewer absences than workers with permanent agreements.

A parallel and growing literature since the middle of 90s, focuses on the employers' provided accommodations and their role in improving the employment of people with disabilities. Burkhauser et al. (1995) are among the first exploring in detail that part of the "Americans with Disabilities Act" (1990) regarding the duties of employers to provide a reasonable placement for workers with disabilities. In particular, it is shown that workplace adjustments are as important as the individual's expectations with respect to the replacement rate of the Social Security Disability Insurance. In a similar vein, Burkhauser et al. (1999) test the importance of these two vectors on the timing of applications for disability benefits, showing how they act in opposite directions. Interestingly, Campolieti (2005) argues that only certain types of arrangements, like a flexible working

schedule or modified workplaces, are associated with a significant increase of employment duration. By focussing on women treated for breast cancer pathologies, Neumark et al. (2015) brings additional evidence on how the type of workplace arrangements can also matter: some types of accommodations such as "assistance with rehabilitative services" bring positive spillovers to women labour supply, while some others like a shorter workday, schedule change etc. push instead negative effects. A recent paper by Hill et.al (2016) states two main points: first of all, there are personality traits such as assertiveness and open communication that are highly predictive of the likeliness to receive a new placement by employers, suggesting how employers can sometimes not even be aware of those needing some workplace arrangements. Moreover, they also find that if employer accommodation rates would have increased, disabled workers would be significantly more likely to delay labour force exit for up to two years. Finally, Anand et al. (2017), by using a sample of people with disabilities who applied for vocational rehabilitation services in three American states, first assess how one-third of reported difficulties (i.e. lack of transportation and an inaccessible workplace) could be potentially addressed by workplace accommodations. Even more interesting are the differences reported by demographic characteristics in perceived barriers. In particular, workplace accessibility is perceived as an employment barrier by those who have lower levels of educational attainment, in poor health and report a physical disability.

Upon the occurrence of an acute health shock, how a *prolonged absence* can affect individuals' working career? Is there sufficient protection against dismissal? Besides the administrative information about health shocks - a significant advantage compares to the previous analyses - the paper approaches a novel research question where many institutional (and not) factors come into play. Dismissal, loss of employability and/or working opportunities could increase the risk of exit from the labour contract, while appropriate on-the-job accommodations could help unhealthy workers to deal with the remaining working life. That is why the institutional setting is extremely important in this type of analysis and thus, throughout the following Section, few remarks on the Italian institutional context will be provided.

3 Institutional Framework

As outlined in the previous sections, Italy is undoubtedly among the countries with the strictest employment protection legislation (EPL) (Scoppa et al. (2014)). However, when asking the main reason for leaving the last job, Figure A2 in Appendix depicts notable differences in answers among blue and white-collar employees. This suggests how some types of workers - and jobs - are more vulnerable than others when health deterioration occurs. In this section, key features of the Italian system will be briefly revised: from the baseline regulation concerning sickness insurance to the employment guarantees offered to people with and without certified disabilities, up to the

employer's costs of running "justified dismissals"[3].

3.1 Sickness Insurance in Italy

A crucial aspect of the Italian sickness insurance system, which is similar to that of most European countries, is that both the public insurer and the private employer are key players: both of them, according to different percentages, must compensate a worker's earnings during his/her absence due to illness. This aspect is crucial within this context of analysis. During the first three days at home, the employer must pay the full wage to sick workers. From the 4th to the 20th days, conditional on a physician certificate (usually provided by the general practitioner (GP)), half of the usual[4] earnings are paid as sickness benefits by the Social Security Institute. From the 21st day up to a maximum of six months, the share rises to two thirds. The majority of the Italian Labour Collective Agreements entrust the employers to cover the remaining part of earnings through the so-called sickness pay [5]. In light of this, when full replacement of earnings is granted to employees, monitoring their real health conditions becomes an important tool. Home visits can be required by both the public insurer and the employer; the latter must pay about 60 euros per visit. As stressed by Biscardo et al. (2019) the employer has multiple reasons for administering home visits: first of all, he is personally involved in payments together with the social security institute; second, he is also encouraged by the increasing organisational costs, especially in cases of long-term absences.
According to the baseline regulation, throughout the period of sick leave coverage - called "comporto" in Italian - the ill worker keeps the rights to his job. The Italian laws, together with each specific national collective agreement, defines the maximum period of coverage. Once the time is expired and the worker is still out of work, the employer can theoretically proceed with a lawful dismissal. In most cases, sick workers can extend the period of absence asking for the available vacations or unpaid "time-off work". Despite these guidelines, exceeding the comporto does not automatically lead to a lay-off as the existence of a justified reason must always be verified.

3.2 Illness and job retention rights

The Italian rules concerning the employment protection of people with long-lasting illness or reduced working capabilities, i.e. individuals unable to perform their previous job tasks, are rather unclear. Where legislative voids arise, erroneous or subjective implementations arise first, and then potential negative consequences not only for sick employees may occur: an unclear definition of

[3]Under art. 2119 C.C., 'just cause', in broad terms, requires very grave conduct which, when evaluated both subjectively and objectively, constitutes a serious and irremediable reason that prevents the parties to continue the employment relationship even on an interim basis. Whether such a breach has occurred would normally have to be determined ultimately by a court, taking all relevant factors into account.

[4]Sickness benefits are computed on the earnings received by the worker during the four past weeks prior to the onset of the disease

[5]As defined by a policy report of the European Commission (Spasova et al. (2016)) "sick pay is the continued, time-limited, payment of the worker's salary by the employer during a period of sickness", while "sickness benefits are provided by the social protection system and are paid as a fixed rate of previous earnings, or a flat-rate amount"

the employers' duties regarding professional integration and reintegration of unhealthy workers may have damaging results on the whole system.

Even before the 90s, the awareness of discriminatory situations on the labour market deriving from health conditions and disabilities was widespread. According to the Article 15 of the Charter of Worker's Rights, any actions such as layoffs, discriminatory assignment of tasks and qualifications, transfers, disciplinary sanctions are null when driven by political, religious, language, sex or disability discriminations. The law 104/92 settled a comprehensive institutional framework to promote the assistance, social integration and rights of persons with disabilities. However, the first attempt to exclude "illnesses" among the reasons for a justified dismissal comes from the Law 68/1999: more precisely, Article 4 refers to disabled workers who become unable to perform their previous job tasks; in this case, they must be reallocated inside the firm without losing the previous economic power. If workplace adjustments are unavailable, they are driven toward different companies where their remaining working capabilities can be better used. Despite the initial purposes and all the refinements defined by subsequent laws (D.L. 216/2003 following the EU directive 2000/78/EC), their effectiveness is prevented by regional implementations. Each region, and often each province, independently manages how these directives must be performed and how to coordinate all related activities. The lack of a clear and univocal national guideline makes it difficult to run any functional employment support initiative.

As additional consideration, the distinction between people with and without certified disabilities is crucial for their application. While disabled individuals are strongly protected against an unlawful dismissal this is not the case for others, who are instead formally subjected to the same rules as the healthy workers. Thus, a more targeted legislation addressed to people with, for example, chronic conditions is still missing. Some exceptions are the following: cancer patients can request a switch from a full-time to a part-time contract (Art.46 D.L. 276/2003), or public-sector employees with cancers can also ask to work from home ("Circolare" 30 April 2009). A step forward in increasing the employment protection of people with reduced working capabilities is the legislative decree 81/2008. According to Article 41, "in case of an absence due to ill health lasting more than sixty consecutive days, it is necessary to check the sustainability of the worker to perform her/his task by a medical examination". Moreover, the next paragraph (the number 42) states the employer must assign the worker to a different, but equivalent, task; if this is not possible, the assignment to a lower duty is also allowed ensuring the previous level of income. In general, the dismissal of an employee with reduced capabilities resulting from a chronic disease is only possible when the employer is unable to find alternative job tasks which are suitable to worker's health conditions whilst always ensuring the good performances for the company. As it is often the case, when room is left for interpretation, a judge must decide how each unique situation must be dealt with. However, since such legal procedures are often costly, the employer's trade-off between the costs of workplace

adaptation and the costs of dismissal is crucial and must always be well considered.

3.3 EPL and firm's costs

The baseline regulation on sickness insurance comes beside the employment protection legislation (EPL) concerning an "unlawful dismissal". The idea of just cause has been extensively defined by the Italian legislation starting from 1970; however, significant differences between small or large firms persist[6]. According to Article 18 Law 300/1970, part of Charter of Worker's Rights, job reintegration's rules are valid when a firm has more than fifteen employees, or six in the case of agricultural firms. In particular, when the dismissal is judged "unfair" the worker is allowed to receive the following payments: a) all the foregone earnings from the period between dismissal and judgement; b) the worker can also decide to either receive an extra financial compensation (corresponding to 15 monthly payments), or to be reinstated inside the firm (Scoppa et al. (2014)). On top of this, the employer must pay all the legal costs together with the penalty for the delayed payment of social security contributions. Thus, large firms face the risk of a costly trial with uncertain outcomes whenever firing a worker becomes necessary (Ichino and Riphahn, 2005). The Charter of Worker's Rights did not mention firms with less than fifteen employees. Despite a subsequent laws extended the criterion of "just cause" to all type of firms, irrespective of their dimension, different regimes of sanctioning are still in place: the employers of small firms may chose between the reintegration of the worker or the payment of a financial compensation ranging between 2.5 and 6 months. In light of this, the incentives for individuals and firms to behave differently according to the firm's dimension are relevant.

4 Data and Descriptive Statistics

4.1 Dataset and sample selection

The research is based on WHIP&Health, an Italian administrative dataset where both health and work histories are collected over time. The baseline population is characterised by a 7% random sample drawn from the *Work History Italian Panel* (WHIP) including a rich set of individual and firm-level characteristics between 1990 and 2012; neither the public nor the agricultural sectors are included. Gender, age, region of birth, area and region of work, the initial and final date of each employment spell, labour income and most importantly, all sickness episodes (paid and unpaid), are among the available information. In addition, information on retirement and other forms of social security benefits (invalidity, unemployment benefits etc.), are also observed. Individual's

[6]Legal safeguards have been reduced since 2011. The 'Fornero-Monti' reform of employment, which came into force in July 2012, rewrote in total article 18 of the Workers' Statute, providing different regulations for different types of dismissal. Its most relevant novelty concerns the possibility for a firm with more than 15 employees to dismiss workers for economic reasons. In this type of dismissal, the employee cannot claim his job back and has only right to an indemnity ranging from 12 to 24 months of salary, the sum being decided by a court. The Fornero-Monti reform thus lessened the restrictions to firing In Italy significantly.

health characteristics are linked with this "employer-employee" database: in particular, the details of all hospitalisations coming from the regional hospital discharge registers (Schede di Dimissione Ospedaliera, SDO) and provided by the Italian Ministry of Health, are collected between 2001 and 2014. The main variables are the primary diagnosis (defined according to the ICD-IX codes) and the length of reference hospitalisation. Despite unavailable, the full version of the dataset includes work injuries and professional diseases recorded between 1994 and 2010 by the National Work Injuries Insurance Administration (INAIL).

The target population is characterised by male workers aged between 18 and 64 years old, who were hospitalised for an acute form of cardiovascular shock - not resulting in death - in a year between 2003 and 2005. More precisely, myocardial infarction and other forms of coronary heart diseases and strokes have been selected (the details of ICD-IX codes is available in Appendix, Table A1). Although no additional restrictions are imposed from the reference hospitalisation onwards, according to the sample selection criteria it is the first hospital admission observed (for a CVD shock) since two years. These conditions become essential requirements for our identification strategy due to the possible endogeneity concerns arising from the main independent variable, i.e. the total number of weeks in sick leave at time $\bar{t}$ with a specific employer (the reference labour contract[7]). The analysis takes advantage of the severity of CVD shocks[8] to control - to some extent - the omitted information related to individuals' attitudes, preferences and unknown health that might be reflected on sick leave as well as on the outcome variable, and thus resulting in biased estimates. On this background, the experience of the first CVD shock since two years allows to better circumscribe an individual's health status and its link with sick leave in a specific point in time.

Furthermore, to reach a homogeneous group of people, both in terms of socio-economic characteristics and institutional framework[9], we only consider blue-collar workers at the time of the reference CVD hospitalisation. Finally, in order to increase both the internal and the external validity of our results, we explore blue-collar workers with *permanent contracts*[10]: besides the constraints in terms of sickness benefit's eligibility and period of coverage, possible findings among permanent employees would alert the policymaker about the adequacy of the employment protection legislation. Indeed, people with permanent contracts are (or, should be) more protected against firing: although specific laws have been implemented in Italy over time in order to prevent people with reduced working-capabilities from involuntary job losses, the limited incentives of the employers may pressure individual working careers in opposite directions.

[7]The reference labour contract is the working-spell at the time the selected CVD shock occurs. In the case of multiple and contemporary jobs, the longest is selected.

[8]CVD shocks are among the leading causes of death in developed countries, including Italy. For men in particular, CVDs represent the most common cause of death under 65 years old (31%) in Europe (compared to about 22% of deaths related to cancer). For women aged below 65 years old, they are the second largest cause of death (26%), after cancer (35%).

[9]Sick leave records are available in the Social Security archives for all blue-collar workers, but only to few categories of white-collars, i.e. those working in the sectors of "Wholesale and Retail Trade" and "Hotels and Restaurants".

[10]By doing this, only 6% of those in the original sample is excluded because they are fixed-term employees.

The final sample is given by 1354 individuals[11]. Figure 1 helps to clarify both the overall structure of the dataset and the aforementioned sampling procedure: as is evident from the picture, the selected health-event (red cross) can occur in a month between January 2003 and December 2005; since then, the residual length of the reference job spell is measured. Job-interruption can either occur before the end of the observational window (December 2012) or after. In the latter case, they are called right-censored observations. The way in which the residual job-tenure is computed will be better explained in Section 1.5.2.

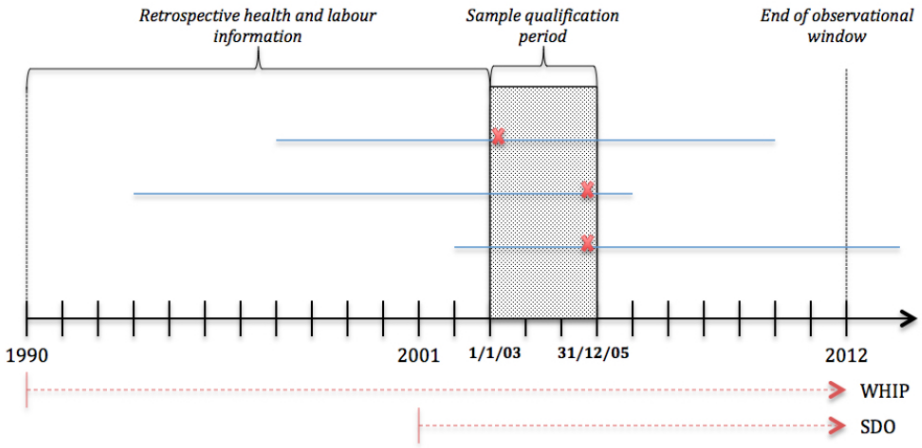

Figure 1: Dataset structure and sampling procedure
Notes: "SDO" refers to the Hospital Discharge Register (*Schede di Dimissione Ospedaliera*, in Italian)

4.2 Variables

The administrative nature of WHIP&Health limits the collection of many demographic characteristics such as the level of education, marital status, or other important information on individual risk behaviours, commonly available in survey data. However, in this case a strong effort has been placed to build a wide set of control variables. Many retrospectives information are available to describe both individuals' health and labour characteristics up to fifteen years prior to the shock (Table 1). The following paragraphs together with an extensive discussion on the potential endogeneity issues arising from sick leave (Section 4.3), aim to further motivate our identification approach.

Current Health Characteristics

Knowing the 'real' health status of people in sickness absence is one of the main challenges not only for the employers but also for researchers aiming to describe the relationship between this measure and various of labour outcomes. Thanks to the sample selection applied - all workers were

[11] In order to limit the misleading effect of extreme outliers, the 1st and the 99th percentiles are also dropped.

hit by an acute form of CVD shock in a given year - it is plausible to consider as severe their current health status. The length of the reference hospitalisation ($days\_cvd\_hosp$) is also available: this additional information - reasonably included in sickness absence and thus, redundant in the main model specification - will be subsequently exploited as an alternative, more objective, 'starting point' when defining the residual job-tenure (i.e. the outcome variable). Besides, when additional hospitalisations have been observed over the year $\bar{t}$, for both cardiovascular and other types of diseases, further covariates help to describe workers' current health conditions. In order to prevent the dimension of the dataset, these situations are taken into account by adding two main variables to the model: the total number of days in hospital for other types of diseases[12] ($days\_others\_\bar{t}$), and the total number of days in hospitals for additional (but subsequent) CVD shocks ($days\_cvd\_\bar{t}$).

Past Health Characteristics

Current health status possibly reflects heterogeneities in past health conditions. Despite the lack of information on risky behaviours, other covariates are useful to this purpose. As it is clear from Figure 1, hospital discharge records start to be collected from 2001 onwards, while labour archives go back to 1990. The structure of WHIP&Health together with the sample selection applied is important to understand the meaning of the following variables: all of them are 'cumulated' up to the year before the reference CVD shock. The variable "$days\_other\_cum$" represents the total number of days in hospitals for illnesses - other than CVD shocks - collected from 2001 up to $\bar{t}-1$. Instead, by referring to the total number of days spent in hospitals for previous CVD shocks up to $\bar{t}-1$ ($days\_cvd\_cum$), it counts all the events that happened two years before. Among the variables built from the labour archives, the total number of weeks in sick leave up to $\bar{t}-1(sick\_leave\_cum)$ offers some insights about their past health conditions. As before, they are cumulated up to $\bar{t}-1$ and consider all the previous jobs, not only to the reference one. Finally, conditional on having a certified level of disability (a reduction of working capabilities of at least 77% must be diagnosed), the Italian social security system allows individuals to receive an ordinary invalidity benefit (OIB) - different from the disability pension - while working. Therefore, aiming to further highlight past health conditions, the variable "$inv\_benefit\_cum$" is also included.

Current Job-related Characteristics

A broad variety of job-specific characteristics are available. Besides common information such as labour income, the type of contract or the area of job etc., the dimension of the firm and the details of the starting and ending date of the reference working spell are relevant for our purposes.

[12]The label "other" refers to aggregate information covering all the possible reasons why an individual might be admitted to the hospital: the inability of distinguishing among their severity is an issue of the dataset.

The variable on the firm's dimension ($firm\_015$) has been defined according to the number of employees. In particular, two reasons are behind the threshold of fifteen employees. On the one hand, bigger firms are often associated with higher levels of job-protection; thus, pushing them to increase unjustified sick leave. On the other hand, bigger firms can more easily adjust job-tasks in case of health-related limitations, allowing a longer labour market activity. The starting date of the reference labour contract together with the month the selected CVD shock, are useful to compute another piece of the story: the seniority of each blue-collar worker with that specific employer up to the time of the shock ($m\_seniority$). Even those who started the job one month before the shock are collected: individuals with a shorter experience, and in particular, with limited seniorities could be treated differently.

Past Labour Characteristics

By using the retrospective WHIP archives, it is possible to collect extensive information on past working histories. The idea is to capture that part of an individual's labour market attachment that makes their effect on heterogeneous working careers. Among them, the total number of years the person has been observed as either employee, a self-employed or atypical worker up to $\bar{t} - 1$. Together with the number of job-spells as employee ($nemployee\_cum$), we aim to describe how long and unstable the career of a person could have been. In a similar vein, we retrieve the variables representing how many times he received an unemployment benefit in the past (proxy of unemployment spells) and the cumulated number of weeks in "cassa integrazione guadagni"(CIG), a partial insurance against unemployment[13] ($nunempl\_cum$ and $ever\_cig$). Finally, additional covariates such as having experienced (or not) self-employed activities in the past or atypical jobs are also included ($ever\_selfempl$ and $ever\_atypical$).

[13]This is an integration or substitution of earnings when working activity has been reduced or suspended due to transitory (difficult) situations. It is thought for specific types of industries (typically manufacturing and construction)

Table 1: Variables name and definition

Variable	Definition
Demographic characteristics	
age	Age at the time of the reference CVD hospitalisation
abirth_north	Area of birth (north)
abirth_center	Area of birth (center)
abirth_south	Area of birth (south)
abirth_islands	Area of birth (islands)
abirth_abroad	Area of birth (abroad)
country_underdev	Equal to 1 if the person comes from an underdeveloped country
Health characteristics at the time of the reference CVD hospitalisation	
sick_leave	Number of weeks in sick leave at the time (year) of the reference CVD shock
sick_leave_paid	Number of paid weeks in sick leave at the time (year) of the reference CVD shock
sick_leave_unpaid	Number of unpaid weeks in sick leave at the time (year) of the reference CVD shock
hosp_cvd_$\bar{t}$	Number of hospitalisations for *other* CVD shocks
days_cvd_$\bar{t}$	Number of days spent in hospitals for *other* CVD shocks
hosp_other_$\bar{t}$	Number of days spent in hospitals for other type of diseases
days_other_$\bar{t}$	Number of days spent in hospitals for other type of diseases
days_cvd_hosp	Days of hospitalisation for the reference CVD hospitalisation
Past Health Characteristics	
hosp_cvd_cum	Equal to 1 if the person ever had a hospitalisation for cardiovascular diseases until $\bar{t}$-1
days_cvd_cum	Number of days spent in hospitals for a cardiovascular shock until $\bar{t}$-1
hosp_other_cum	Equal to 1 if the person ever had a hospitalisation for other diseases until $\bar{t}$-1
days_other_cum	Number of days spent in hospitals for other type of diseases until $\bar{t}$-1
inv_benefit_cum	Equal to 1 if the person ever received ordinary invalidity benefits until $\bar{t}$-1
sick_leave_cum	Number of weeks in sick leave until $\bar{t}$-1
sickleave_paid_cum	Number of paid weeks in sick leave until $\bar{t}$-1
sickleave_unpaid_cum	Number of unpaid weeks in sick leave until $\bar{t}$-1
Current Job Characteristics	
labour_income	Annual earnings
part_time	Equal to 1 if the person is a part-time employee
s_primary	Equal to 1 if the person works in the primary sector of activity
s_secondary	Equal to 1 if the person works in the secondary sector of activity
s_tertiary	Equal to 1 if the person works in the tertiary sector of activity
awork_north	Area of work (north)
awork_center	Area of work (center)
awork_south	Area of work (south and islands)
firm_015	Equal to 1 if the person works in a firm with less or equal than 15 employees
m_seniority	Months of seniority (with the same employer) up to the month of the reference CVD shock
Past Job Characteristics	
work_active_cum	Number of years the person is observed as employee, self-employed or atypical worker until t-1
nemployee_cum	Number of contracts as employee until $\bar{t}$-1
ever_selfempl	Equal to 1 if the person ever worked as self-employed until $\bar{t}$-1
ever_atypical	Equal to 1 if the person ever worked as atypical worker until $\bar{t}$-1
nunempl_cum	Number of unemployment benefits received until $\bar{t}$-1
ever_cig	Equal to 1 if the person ever been in "cassa integrazione guadagni" until $\bar{t}$-1

4.3 Endogeneity of sick leave

The *possible* endogeneity of sickness absence is one of the main concerns often discussed by empirical researchers. Omitted information relating to both individual attitudes and underlying health conditions can lead to biased estimates. This section offers an extensive review of such an empirical issue, discussing the use of sample selection criteria and the full set of controls as a way of dealing with it in the following regression models.

A primary source of omitted information can arise from the *current* health condition, which is inherently latent. Hence, to what extent the number of weeks in sick leave is associated with the post-shock health status becomes essential in this context of analysis. The unknown health

should be negatively correlated with the length of sickness absence and, at the same time, positively correlated with the residual stay in a labour contract. By focusing exclusively on those who have experienced an acute CVD shock in a specific point in time ($\bar{t}$), it is palusible that a sizeable part of the time spent on sick leave during that year strongly correlates with current health conditions. Different factors can support this argument: first of all, the information about sick leave is encoded by the Italian Social Security Institute as the number of entire weeks of absence due to illnesses, meaning that a few days at home taken by the workers (arguably correlated with own preferences) are not accounted for, and thus cannot confound the studied relationship. Here the use of an administrative source is an undeniable advantage with respect to the most common survey data where that measure is typically registered as the "number of days": a daily (and self-reported) evaluation potentially reflects diverse situations, some of them unrelated to the worsened health conditions. Secondly, according to the Italian regulation on sickness insurance, after three days at home, a GP's certificate is required, meaning that for longer time spans a professional doctor guarantees a worker's health conditions. Although some papers in the literature argue the subjective nature of judgements for certification practices among physicians (Askildsen et al. (2005), Markussen et al. (2011)), we claim this is less of an issue when CVD shocks occur. Moreover, the Italian legislation allows external medical visits, required by both the Social Security Institute (INPS) and by private employers, aiming to check the real workers' conditions and thus discourage absenteeism behaviours[14].

Another source of missing information possibly derives from past health conditions. Information on past health events and risky behaviours such as smoking habits, drug use, etc. are valuable when controlling for factors that correlate with, and possibly explain, the current health status. Although this information is often available in survey datasets, it is usually absent in administrative ones. However, WHIP&Health allows observation of more objective health episodes as they come from the national hospital discharge registers. As explained in the previous section, the paper extensively explores all possible information contained in both components of the dataset - WHIP and SDO archives - to enrich the set of control characteristics.

Eventually, a complete discussion of the possible endogeneity of sick leave cannot avoid mentioning the role of individual preferences and personal attitudes. For instance, low effort and low risk-averse workers may take advantage of generous sickness benefits schemes to extend the absence period beyond their real needs. This behaviour is the so-called absenteeism. On the contrary, as increasingly stressed in the literature, individuals at the left tail of the income distribution tend to shorten their sick leave periods at the expense of their health and of proper recovery as they fear losing their job and are also more vulnerable to income drops. This is commonly known as presenteeism behaviour, i.e. working while sick. Still, some considerations come in favour of our approach: again, the selection of CVD diseases enables us to focus on health shocks for which a doctor is asked to

[14]After a home visit, the external doctor declares whether or not the employee is fit to return to work within three days.

evaluate and recommend a proper recovery period. Additionally, behaviours like absenteeism or presenteeism, or factors such as individual discretionally (which mainly occur for less severe illness), are of minor importance. Furthermore, as mentioned before, institutional features such as "home visits" work as a strong disincentive against opportunistic behaviours. As proved by Biscardo et al. (2019), as long as the public insurer is not supported by specific algorithms[15], the private employer undoubtedly has an informative advantage when choosing which workers to visit, making home visits even more effective. As a final remark, personal attitudes and preferences will be further taken into account to the extent they correlate with observed characteristics. The Work History Italian Panel (WHIP) suits this task as it contains lots of information about their previous working history, which arguably correlates with and accounts for an employee's general attitude to the work. For instance, workers with stable and long career paths may be tempted to extend the period at home while sick and, on the contrary, young workers with unstable jobs could hurry up the recovery period and get back to work as soon as possible in order to show off their attitude to the employer. Variables such has the length of observed working career since 1990 ($work\_active\_cum$), the total number of unemployment spells ($nunempl\_cum$) or the number of different jobs as an employee up to $\bar{t}-1$ ($nemployee\_cum$), are only a few examples of useful predictors for these individual behaviours and attitudes, and will therefore be included in the following regression models.

[15]Since March 2011, the selection of workers' sick leave to minor is addressed by the public insurer by a "data mining software".

4.4 Descriptive Statistics

Before looking the distribution of the available covariates, a preliminary exploration of the studied event - the time until job-interruption occurs - is needed. Basic statistics on survival times show that for 24% of people we do not observe the expiring date of the reference labour contract; the first quartile of subjects survive in that job less than one year (8.86 months), while half of them less than 3 years (34.9 months). Figure 2 shows a non-parametric estimate, called Kaplan-Meier estimator[16], of the probability of survival past time t in the reference labour contract. The measure of sick leave in $\bar{t}$ has been divided in three different groups[17] according to its distribution: the blue line refers to those who spent between 1 and 6 weeks at home, the red line represents people who make between 7 and 18 weeks, while the green line characterises those with more than 19 weeks of absence. Unsurprisingly all the curves follow decreasing trends: month-after-month, the number of those who get out from that job increases, while those remaining decrease. Interestingly, clear differences appear in survival rates over the first six years after the return-to-work: individuals who did between 1 and 6 weeks in sick leave are more likely to survive in the reference contract beyond each point in time, as opposed to those with more than 19 weeks face a huge drop since the beginning. From the sixth year onwards, blue-collar workers in the middle and those in the lowest part of the distribution experience similar patterns. Besides the graphical evidence, it is possible to formally test the hypothesis for the equality of survivor functions across groups by using two different tests: the Log-rank[18] and Peto-Peto-Prentice[19] tests. Under the *null* hypothesis all survival curves are the same: in both cases, the equality of survival functions is rejected at 1% confidence level.

An extensive overview of the covariates' distribution is available in Table 2. In light of the type of health shock considered, the current number of weeks in sick leave is not surprising: on average, they take more than 12 weeks to recover (approximately 3 months); most of the time is paid by sickness benefits, while a smaller fraction is unpaid[20]. Similar reasons can be mentioned in order to justify the average age (50 years old) of the selected individuals; although the value is pretty high, it is in line with the general national and international statistics[21] (as age increases, also their incidence grows). With respect to the reference hospitalisation, they spend approximately one week inside the

[16]The Kaplan-Meier estimator is defined as follows:

$$\hat{S}(t) = \prod_{j|t_j \leq t} (\frac{n_j - d_j}{n_j})$$

where n_j is the number of individuals at risk at time t_j and d_j is the number of observed failures at time t_j. The distance to the failure event (i.e. job-interruption) is computed in months, thus the estimator performs the evaluation month-after-months.

[17]People in the second and third quartile of the distribution have been grouped because showing similar trends in terms of survival times.

[18]The Log-Rank test is a large-sample chi-square test that uses as its test criterion a statistic that provides an overall comparison of the KM curves being compared (Kleinbaum et al.(2005)

[19]All the other types of tests are variations of the log-rank test statistic and are derived by applying different weights at the jth failure time. Peto test weights the jth failure time by the survival estimate calculated by considering all groups combined (Kleinbaum et al. (2005)

[20]This second option is allowed when the "comporto" is exceeded.

[21]http://www.salute.gov.it/imgs/C_17_navigazioneSecondariaRelazione_1_listaCapitoli_capitoli ItemName_1_scarica.pdf

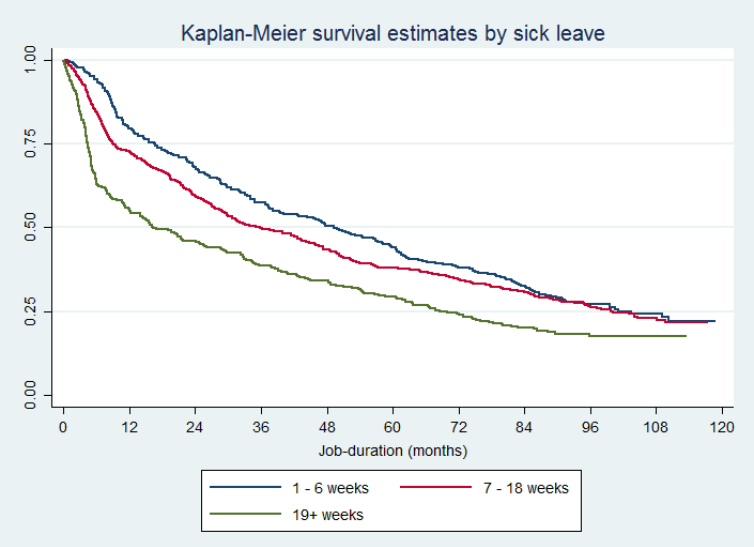

Figure 2: Kaplan-Meier survival estimates *by sick leave groups* in $\bar{t}$
Source: WHIP&Health. Notes: Log-rank and Peto-Peto-Prentice test of equality of survivor functions (χ^2(Prob)) have been performed: 26.22 (0.000) and 45.25 (0.000), respectively.

hospital (7.4 days), rarely more than 9 days (75th percentile). This is reasonable considering that the most severe cases are followed by death or the patients are moved to specialised structures. Figure 3 depicts the distribution of sick leave and the number of days of hospitalisation, respectively. The statistics on the average number of additional events (and days spent in hospitals) for cardiovascular diseases are not exhaustive. Indeed, although the average number of hospitalisations is close to zero (0.4), a sizeable part of people (almost 30%) have at least one subsequent and very close new event[22]. Even "other types of health events" regularly occur, counting more than 30.1%. In this case, the number of days in hospitals is higher than CVD episodes, on average, 5.28 against 2.36. The reason of that can be found in the broad variety of diseases the label "other types of health events" covers: either a hip fracture or malign cancers can fall into this category.

Lagged health characteristics should be critically discussed thinking about our sample selection: as extensively stressed before, none of the selected subjects has any hospitalisation for CVD shocks during the two previous years. Thus, unsurprisingly, both the average of cumulated episodes and the days in hospitals are pretty low. Slightly different is the case of the other type of hospitalisations: driven by those in the last quartile having experienced more than one hospitalisation, the average number of days is 2.61. By comparing the values referring to $\bar{t}$ and $\bar{t}-1$, the latter is substantially lower than the former, suggesting how the general health conditions of these people decreased significantly by the time of the reference shock. With respect to the number of past weeks in sick leave up until $(\bar{t}-1)$, on average, 19 entire weeks in sick leave are counted, while the median individual reports 10 weeks. Reasonably many factors can be reflected in these values. For instance, a 35 years old blue-collar worker could have half of the weeks either because he is younger

[22]According to the sample selection, people are allowed to experience new CVD immediately after the reference one.

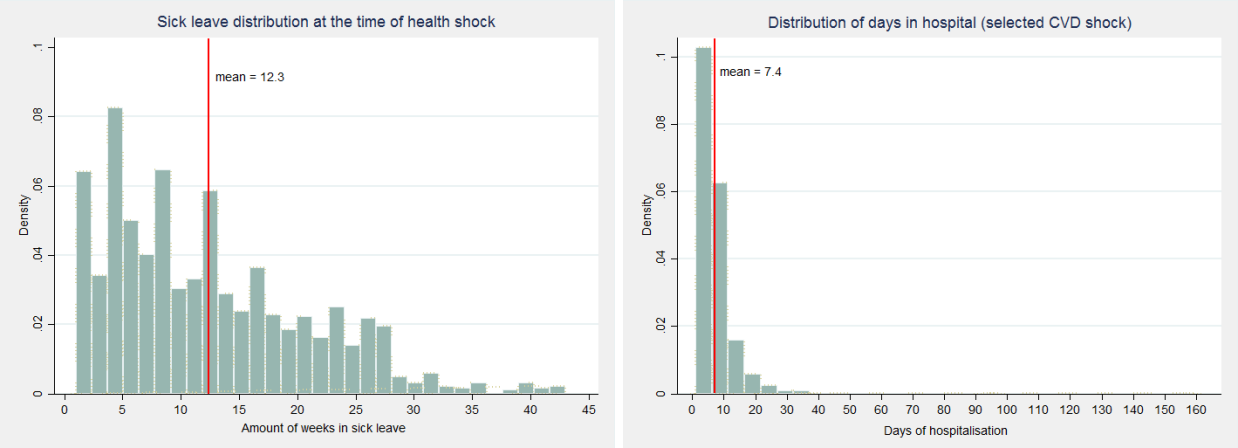

Figure 3: *Distribution of sick leave in $\bar{t}$ and length (days) of the reference hospitalisation*
Source: WHIP&Health. Notes: the vertical red line refers the average values

and healthier, or as a consequence of a shorter and more discontinuous working career characterised by lots of years of inactivity - especially as an employee - and unemployment spells.

With respect to the current job characteristics, the average seniority up to the time of the shock is 106 months, slightly less than 9 years. Most of them are full-time workers employed in Northern regions by secondary sector firms. Their gross labour income rarely exceeds 30.000 euros and more interestingly, 30% of them is employed by firms with less than fifteen employees. When looking at their past labour characteristics the first thing to stress is that, on average, they entered INPS archives more than 11 years before; since the 50th percentile they are already observed through 13 years. Only the 11% have had at least one period as self-employed in the past, and very few (2%) also ran atypical jobs. The number of past episodes in unemployment is rather limited, although 38% of them have experienced at least a period in "cassa integrazione guadagni".

Table 2: Descriptive Statistics

	Mean	SD	Min	Max	p50
Demographic characteristics					
age	50.47	7.45	22	64	52
abirth_north	0.271	0.45	0	1	0
abirth_centre	0.139	0.35	0	1	0
abirth_south	0.345	0.48	0	1	0
abirth_islands	0.143	0.35	0	1	0
abirth_abroad	0.102	0.30	0	1	0
country_underdev	0.094	0.29	0	1	0
Current Health Characteristics					
days_cvd_hosp*	7.400	7.56	1	162	6
sick_leave	12.30	8.41	1	43	10
sick_leave_paid	10.98	8.29	0	43	9
sick_leave_unpaid	1.205	3.68	0	38	0
hosp_cvd_$\bar{t}$	0.389	0.70	0	6	0
days_cvd_$\bar{t}$	2.362	6.25	0	96	0
hosp_other_$\bar{t}$	0.493	0.91	0	7	0
days_other_$\bar{t}$	5.275	17.6	0	212	0
Past health characteristics					
sick_leave_cum	19.32	25.7	0	272	10
sickleave_paid_cum	17.75	24.1	0	266	10
sickleave_unpaid_cum	1.567	5.86	0	103	0
inv_benefit_cum	0.067	0.25	0	1	0
hosp_other_cum	0.517	1.04	0	8	0
days_other_cum	2.612	7.05	0	82	0
hosp_cvd_cum	0.030	0.24	0	4	0
days_cvd_cum	0.162	1.35	0	18	0
Current Job Characteristics					
labour_income	23056	9377	695.4	137943	21934
part_time	0.058	0.23	0	1	0
m_seniority	106.4	86.9	0	254.6	73.1
s_primary	0.067	0.25	0	1	0
s_secondary	0.691	0.46	0	1	1
s_tertiary	0.242	0.43	0	1	0
awork_north	0.500	0.5	0	1	0.5
awork_centre	0.188	0.39	0	1	0
awork_south&islands	0.312	0.46	0	1	0
firm_015	0.309	0.46	0	1	0
Past Job Characteristics					
work_active_cum	11.82	3.62	1	15	13
nemployee_cum	12.32	4.49	1	27	14
ever_selfempl	0.106	0.31	0	1	0
ever_atypical	0.023	0.15	0	1	0
nunempl_cum	0.423	1.39	0	11	0
ever_cig	0.384	0.49	0	1	0

Source: WHIP&Health. **Notes**: the variable denoted with (*) is specific of the *reference* CVD hospitalisation.

5 Econometric modelling

5.1 Duration analysis and Cox Proportional Hazard Model

The aim of this paper is to study the relationship between the length of sickness absence under severe health conditions and the residual job-tenure. As extensively discussed in the previous sections, opposite mechanisms can differently influence this relationship, especially when permanent jobs are considered. Thanks to the detailed information available such as the starting and ending dates of the labour contract and the month of the CVD shock, duration models become a useful tool of analysis.

Duration models are based on two complementary elements: the *survivor function* ($S(t)$) is the probability of surviving beyond time t (where t is the elapsed time since the first entry into the risky set)[23]; while the *hazard function* ($h(t)$), also known as the conditional failure rate, gives the instantaneous potential for failing at time t per unit of time, given the survival up to time t[24]. Rather than a probability, the hazard function is a rate, and thus it is constrained to range between zero and infinity. According to the Cox PH regression model (Cox (1972)), a semi-parametric approach, the hazard rate can be defined as:

$$h(t, \mathbf{X}) = h_0(t) e^{\sum_{i=1}^{p} \beta_i X_i} \quad (1)$$

where $h_0(t)$ is the baseline hazard function and X_i is a vector of time-independent covariates, either continuous or dummies, measuring individuals' health and labour status (i) at the time of the shock ($\bar{t}$) and (ii) up to the year before (i.e. $\bar{t} - 1$). Our main interest is the predictor "sick_leave", i.e. a continuous variable capturing the total number of weeks in sick leave during the year of the reference CVD shock. As it is clear from the equation (1), while the baseline hazard function depends on the time component, the covariates do not. An alternative version of Cox PH model, the so-called extended Cox model, allows to consider time-dependent covariates. Despite the potential advantages, we consider this approach not suitable in our context: moving the covariates over time instead of keeping them fixed at $\bar{t}$, would have further increased the risk of including the confounding effect of individual attitudes and preferences.

In general, the knowledge of how the risk of the studied event moves over time - in this case, the time until job-interruption occurs - guides the choice between parametric or semi-parametric approaches. When the underlying functional form of $h_0(t)$ is known (exponential, Weibull, lognormal etc.), parametric estimates are preferred. Otherwise, to avoid misspecification problems, semi-parametric methods are the best choice when the functional form is unknown: it has been shown how

[23] S(t) = 1 - F(t) = Pr(T>t) where F(t) is the cumulative distribution function

[24]
$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t + \Delta t > T > t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}$$
where f(t) is the density function. In few steps we could easily demonstrate the relationship between the survivor and hazard functions.

these methods closely approximate the results of the *correct* parametrization. Indeed, this paper takes advantage of the flexibility of Cox proportional hazard models to estimate the relationship between the hazard function and a set of explanatory variables without "making assumptions about the shape of the baseline hazard over time" (Cleves et al. (2010)). In other words, the baseline hazard function is simply left non-estimated thanks to an alternative method proposed by Cox (1972) - hence the name of the approach - where a Partial Likelihood (PL) function substitutes the common Maximum Likelihood estimator (ML). The PL estimator essentially orders *events*, not persons, and thus, instead of considering the probabilities of all units, it takes into account only the probabilities retained from those who are observed to fail (Jenkins (2005)) [25]. Parametric estimates will be run in section 1.7 as an additional robustness check of the baseline results.

In general, whatever it is the shape of $h_0(t)$, the baseline hazard is assumed to be same for every unit. Thus, the hazard ratio (HR) between two different individuals, whose characteristics are represented by the vectors of covariates $\mathbf{X}^*$ and $\mathbf{X}$ respectively, can be written as:

$$\hat{HR} = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})} \quad (2)$$

By substituting the equation (1), we get an additional expression which only depends on the vectors of covariates $\mathbf{X}^*$ and $\mathbf{X}$, while the baseline hazard cancel out:

$$= \frac{\hat{h}_0(t) exp[\sum \hat{\beta}_i X_i^*]}{\hat{h}_0(t) exp[\sum \hat{\beta}_i X_i]}$$

$$= exp[\sum_{i=1}^{p} \hat{\beta}_i (X_i^* - X_i)] \quad (3)$$

The proportional hazard assumption (PH) requires that the an individual's hazard is proportional to the hazard of another individual, and the ratio does not depend on time. In Section 1.7 the validity of the PH assumption will be tested by running appropriate checks; moreover, as additional sensitivity analysis, a stratified Cox PH model will be also performed. Rather than assuming that everyone face the same baseline hazard, this alternative specification allows $h_0(t)$ to differ among different groups according to the predictors not satisfying the PH assumption. As consequence, the hazard function can be slightly modified:

$$h_s(t, \mathbf{X}) = h_{0s}(t) e^{\sum_{i=1}^{p} \beta_i X_i} \quad s = 1, ..., S$$

where S is the total number of strata.

5.2 Basic setup: event, risk period and censoring definition

The main feature of duration analysis is to model the instantaneous probability of a transition from one state to another one: the occurrence of a specific event sets that transition.

[25] We remind to Jenkins (2005) for an extensive explanation about the approach and survival analysis in general.

The event, or *failure*, is now defined as the job-interruption (the exit from the reference labour contract), while the time since the individual can potentially fail is the *risk period* (residual job-tenure). Hence, each point in time (monthly defined) will be characterised by people who exit from the labour contract and subjects for whom the event has not yet occurred.

Upon this setting, a clear definition of the residual job-tenure turns out to be crucial: while the ending point is straightforward, less clear what the starting point is. Although the available data do not allow a precise link of each hospitalisation with the number of subsequent days/weeks at home, thanks to the type of shock considered we are pretty confident that most of the weeks in sick leave during that year are associated to the reference hospitalisation. Therefore, knowing the month in which the reference CVD shock occurs, we place the return-to-work "n-weeks after that point in time", i.e. the number of weeks in sick leave observed in $\bar{t}$. Accordingly, the residual-job

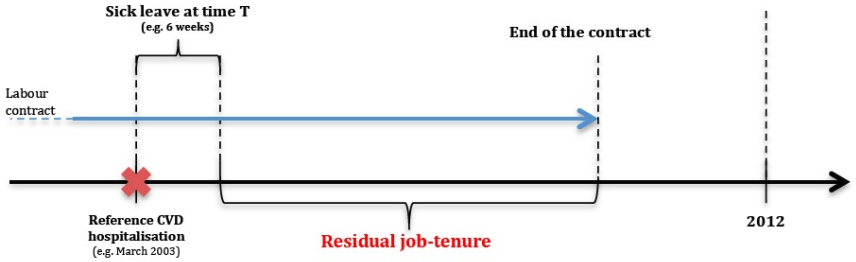

Figure 4: Graphical representation of the general setup

tenure is computed as the distance (in months) between the return-to-work, after the occurrence of the reference CVD shock, and the end of the reference contract. Figure 4 offers a graphical representation of this measure. When the failure event (job-interruption) has not yet occurred by the end of the observational window, right-censored subjects appear. In this study, WHIP archives collect information on the individual working careers up to 2012, in particular December 2012. Among our observations, 329 individuals (24.26% of the total) are still under the same employer. Dealing with censored data is one of the main advantages of using survival analysis; an extension (mainly on the right) of the sample qualification window would have further increased the number of these situations.

As a final remark, it is worth mentioning that when multiple hospitalisations are observed in $\bar{t}$, either for "other types of diseases" or additional (subsequent) CVD shocks, a misleading association between the number of sick leave and the reference hospitalisation may arise. In order to preserve the frequencies, the following Cox PH models will take these situations into account by including specific variables such as the total number of days in hospitals in $\bar{t}$ for CVD shocks or other illnesses; later on, various sensitivity analyses will be also run to test measurement concerns.

6 Empirical Results

The popularity of Cox models relies on a key factor: differing from the parametric approaches, the functional form of the baseline hazard is unknown and thus, left unspecified. Accordingly, the estimated hazard ratios can be interpreted as the instantaneous relative risk of exit from the labour contract - conditional of being survived up to t - given a unitary change in a specific covariate (all the others kept constant[26]). Table 3, where four different specifications of Cox PH models are presented, aims to investigate the role of potential confounders in studying the relationship between the length of sick leave and the survival in a permanent job. Subsequently, some heterogeneities in the effect of extended absences will be also considered in Table 4.

Besides all the available covariates, the main independent variable is the number of weeks in sick leave during the year of the reference CVD hospitalisation. On the one hand, thanks to the selection applied, the length of sick leave can plausibly reflect an individual's time needed to recover. On the other hand, an extended absence (a negative signal in terms of productivity), drives many possible situations: for instance, the employer can either make an effort to find a new accommodation inside the firm or conversely, make additional pressure for an early job-interruption. In light of this, the strictness and the effectiveness of legislative constraints matter. Irrespectively to the model specification, all the values in Table 3 show a negative relationship between the two measures: as the absence increases, so does the instantaneous risk of exit from that contract (about 1.6% for each additional week at home). It is worth noticing how the hazard ratios estimated in Table 3 are all strongly robust across alternative specifications. In other words, they seem insensitive to controls, suggesting how the 'time needed to return-to-work' does not reflect the confounding effect of hidden factors. This evidence, even though incomplete, points in favour of the baseline idea that, thanks to the specific sample selection it is reasonable to consider the number of weeks spent at home the result of a doctor's judgement based on the severity of the shock occurred. For instance, if the relationship between sick leave and the subsequent risk of closing a job is spurious because of workers' presenteeism or absenteeism behaviours, then we would arguably observe a positive correlation between the current and the past-cumulative days spent on sick leave. In that case, controlling for 'past health characteristics' in the model should absorb part of this effect and would attenuate the magnitude and the significance of the main coefficient. Another possibility is when the relationship is spurious because of heterogeneities in employment careers: recently hired employees may have the incentive to reduce absences relative to more senior workers. Even in this case, part of this effect would be captured by variables such as working experience and the seniority within the firm, and the coefficient would be lower when switching from specification 1 to 4.

In detail, the first model includes few essential covariates regarding an individual's health

[26]When X_1 is a dichotomous variable for example, HR is the difference in instantaneous risk among those who have a specific characteristic and those who have not, keeping constant all the others

Table 3: Effect of sick leave duration on job-interruption - *baseline*

	Model 1	Model 2	Model 3	Model 4
sick_leave	1.017***	1.018***	1.015***	1.016***
	(0.004)	(0.004)	(0.005)	(0.005)
Past Health Characteristics	NO	YES	YES	YES
Current Job Characteristics	NO	NO	YES	YES
Past Job Characteristics	NO	NO	NO	YES
N.Obs.	1354	1354	1354	1354
Log pseudolikelihood	-6722.4	-6722.1	-6683.8	-6668.8
Wald chi2	283.0	286.3	369.3	395.2

Source: WHIP&Health. **Notes**: The table reports the *hazard ratios* of four different specifications of Cox PH models. The variable "sick_leave" refers to the whole number of weeks in sick leave at the time (year) of the reference CVD hospitalisation. Robust standard errors have been considered. * p<0.1, ** p<0.05, ***p<0.01

status at the time of the reference CVD shock ($\bar{t}$) and his demographic characteristics. The second specification takes into account the confounding effect of past health conditions; while the third one also includes firm and job's characteristics. The last model instead, considers all the covariates mentioned above plus a set of controls describing individuals' past working life and their labour market attachment: unknown attitudes and preferences for working activity may potentially affect the amount of sick leave as well as the residual job tenure. The full estimates are available in the Appendix (Table A2). Unlike the initial expectations, past health conditions appear slightly powerful - per se - in predicting the risk of exit from that labour contract. On the contrary, some interesting results come out from the current job characteristics: working less hours is reasonably one of the first post-shock workplace adjustments; thus, when switching to part-time is not allowed, job-interruption is the only alternative. Indeed, part-time workers show a lower instantaneous risk of ceasing the reference labour contract than those employed full-time. Moreover, as remarked by Scoppa et al. (2014), notable differences appear among people working in small rather than bigger firms. Individuals employed in large companies, here firms with more than fifteen employees, are doubly advantaged against job-interruption: first of all, firms would face higher costs in case of an unlawful dismissal; secondly, they allow an easier reallocation of tasks when working capabilities reduce. The *Likelihood Ratio* statistic[27] - typically used with ML estimates - is useful to test the significance of the covariates added along the four different specifications. In particular, while no significant improvements have been found between the first two models, past job characteristics together with the current ones, significantly increase the fit of the model. From now on, the last one will be used as the reference one.

Table 4 takes into account potential heterogeneities hidden behind our main independent variable [28]. Due to dimensionality concerns, the use of interaction terms turns out to be particularly suitable in our setting. By including interactions, we allow the effect of our main variable *sick leave* to increase or diminish with the level of an another factor: in particular, if the interaction is

[27] The LS test has been performed on the two models without the option of robust standard errors
[28] The extended version of the regression models are available in Appendix (Table A3)

greater than 1, a positive relationship between the two variables arises; if it is lower than 1, a negative relationship turns out; a value equal to 1 suggests a constant effect. The first type of interaction aims to investigate whether the effect of an equal amount of sick leave differently affect the instantaneous risk of job-interruption among younger and older groups of people [29]. On the one hand, younger individuals are likely to better recover after a CVD shock, keeping their levels of pre and post-shock productivity constant. On the other hand, preferences for leisure and expectations about future working lives can strongly differ among the two groups: while younger blue-collar workers retrain themselves encouraged by a longer remaining working-life, the elderly can easily exit from the labour market through early retirement. Besides, the role played by employers is non-negligible. Some of them can find retraining the youngest more convenient if it brings higher profits in the future. On the contrary, some others might consider the reallocation of the elderly less costly due to past investments. Column 2 of Table 4 shows our findings: surprisingly, no significant differences appear in the effect of sick leave among age groups. Despite negligible, the only exception appears in the second interaction ($sickleave\_age4751$): at a confidence level of 10%, the effect of one more week in sick leave among people aged between 47 and 51 years old, is 2 percentage points lower than the risk of the elderly (the baseline group). Although the available data do not allow to disentangle the potential adjustment channels, it is clear that all individuals experience increasing difficulties in their post-shock working activity. Irrespectively of age, the occurrence of an acute CVD shock has undoubtedly negative consequences on individuals' labour market participation. Whatever it is the underlying reason (from a discriminatory work environment to a missing reallocation of tasks or incentives to retrain), being in a permanent contract does not prevent blue-collars against the economic deprivations followed by a severe health shock.

With the second interaction instead, we aim to partially address the question of whether or not the employers behave differently according to firm dimension. As previously mentioned, companies with a number of employees above or under fifteen units face very different costs when unfair dismissal are stated. As a consequence, the effect of an additional week at home on the instantaneous risk of ceasing the reference job can potentially be higher in smaller than bigger firms[30]. Even in this case the interaction term turns out insignificant and equal to 1, thus the effect does not differ for dimensionality reasons. Despite this, the overall effect of firm dimension is relevant *per se*: working in a small company increases the instantaneous risk of job-interruption of about 43%. Such a huge difference might be the result of many reasons: from the easiest reallocation of the unhealthy workers in big firms to the higher costs of unlawful dismissals. Similar outcomes had already appeared throughout the baseline estimates.

[29] Age group dummies have been defined according to the age distribution. The median value is equal to 52 years old.

[30] Not surprisingly, among the companies with less than 20 employees permanent contracts were 68% of all the type of contracts stipulated in 1998, reaching 77% in 2001; bigger firms register values of 35% and 20%, respectively. (Contini, 2018)

Table 4: Effect of sick leave duration on job-interruption - *Interactions*

	No Interactions (Model 4)	**Interaction 1**	**Interaction 2**
sick_leave	1.016***	1.021**	1.016***
	(0.004)	(0.01)	(0.005)
age1846	0.331***	0.368***	0.331***
	(0.03)	(0.06)	(0.03)
age4751	0.390***	0.500***	0.390***
	(0.04)	(0.08)	(0.04)
age5255	0.690***	0.667**	0.690***
	(0.06)	(0.10)	(0.06)
sickleave_age1846	-	0.991	-
	-	(0.01)	-
sickleave_age4751	-	0.980*	-
	-	(0.01)	-
sickleave_age5255	-	1.002	-
	-	(0.01)	-
firm_015	1.431***	1.431***	1.429***
	(0.11)	(0.112)	(0.177)
sickleave_firm015	-	-	1.000
	-	-	(0.01)
Past Health Characteristics	YES	YES	YES
Current Job Characteristics	YES	YES	YES
Past Job Characteristics	YES	YES	YES
N.Obs.	1354	1354	1354
Log pseudolikelihood	-6668.8	-6666.4	-6668.8
Wald chi2	395.15	399.7	395.8

Source: WHIP&Health. Notes: The table compares the *hazard ratios* of the last model in Table 3 (Model 4) with two additional specifications of Cox PH model: "interaction 1" includes some interaction terms between sick leave and different age groups (people aged between 56 and 64 are the baseline); "interaction 2" includes an interaction between sick leave and the dummy variable for firm dimension (above or under fifteen employees). Robust standard errors have been considered. * p<0.1, ** p<0.05, ***p<0.01

7 Sensitivity Analysis

As extensively explained in the previous sections, the residual job-tenure and thus, the empirical analysis, has been performed by assuming that the number of weeks in sick leave observed in $\bar{t}$ is linked to the reference CVD hospitalisation. Despite the confidence of a strong linkage among them, the first set of sensitivity analysis is primarily employed to loosen possible measurement issues. The second part of this section instead, is devoted to relax the PH assumption, while the last set of analyses offers an overview of the alternative ways in which duration models can be performed.

The association between the reference CVD hospitalisation and sick leave can be imprecise, especially when people experience additional - but subsequent - acute CVD hospitalisations during the same year ($\bar{t}$). Although the awareness that such measurement issue cannot be completely solved due to the existence of bad-health episodes that are not associated with a hospitalisation, we aim to reinforce the validity of our previous findings by first excluding those who had additional hospitalisations for CVD diseases in the same year as the referenced health episode[31]. The results available in column 2 of Table 5 (*sensitivity 1*) are strongly encouraging: the instantaneous risk of ceasing that job is still significant at 1% confidence level and the magnitude is equal to the main

[31]In the case of new CVD shocks, the sample selection ensures they occur strictly after our reference CVD hospitalisation.

result in the previous section (model 4). Despite dimensionality concerns, we further exclude those who had other types of health shocks in $\bar{t}$ besides CVD shocks: the HR turns out insignificant but the negative relationship persists (see Table A4 in Appendix).

Table 5: Effect of sick leave duration on job-interruption - *Sensitivity Checks*

	Baseline (Model 4)	Sensitivity 1	Sensitivity 2
sick_leave	1.016***	1.016***	1.005
	(0.004)	(0.005)	(0.004)
Past Health Characteristics	YES	YES	YES
Current Job Characteristics	YES	YES	YES
Past Job Characteristics	YES	YES	YES
N.Obs.	1354	954	1354
Log pseudolikelihood	-6668.8	-4408.9	-6683.6
Wald chi2	395.15	342.4	398.8

Source: WHIP&Health. Notes: The table compares the *hazard ratios* of the last model in Table 3 (Model 4) with two additional specifications of Cox PH model: "sensitivity 1" excludes those who experienced additional hospitalisations for CVD diseases in the same year as the referenced CVD hospitalisation. "sensitivity 2" performs a Cox PH model considering a residual job-tenure which starts the day after the discharge from the reference CVD hospitalisation. Robust standard errors have been considered. * p<0.1, ** p<0.05, ***p<0.01

As explained before, one of the main elements of duration analysis is 'the time until an event occurs'. Hence, how the distance to this event is computed becomes important. The way how the residual job-tenure has been previously defined, i.e. placing the whole number of weeks in sick leave after the reference CVD shock, kept us safe from an inappropriate measure of the distance: without considering the length of sick leave in $\bar{t}$ the duration model would not be able to disentangle the effect of longer or shorter absences. Accordingly, as explained in Section 1.5.2, the return-to-work had been placed n-weeks after the occurrence of the reference CVD shock. However, the concern of a wrong association between our main independent variable "sick leave" and the reference hospitalisation, remains. Besides the initial approach, we can alternatively test the baseline results by setting a different starting point, and thus a different distance: taking advantage of the number of days of hospitalisation, information which is only available for the reference CVD shock, we place the return-to-work just after the discharge. Therefore, the residual job-tenure is computed as the distance from this new point in time and the job-interruption; the results are available in Table 5 (*sensitivity 2*). Unsurprisingly, our main independent variable turns out insignificant at a confidence level of 5% and 10%. This result confirms what has been stated before; despite its objectiveness, such a distance is unable to adequately capture the role played by an extended period at home on the subsequent risk of leaving the job. Therefore, we claim the baseline results are still rather robust.

Thanks to the previous analyses, the measurement issues arising from the structure and the drawbacks of the dataset can be substantially decreased. We are now going to relax one of the main assumptions of the Cox proportional hazard model, i.e. the hypothesis that everyone faces the same shape of baseline hazard. With this in mind, we claim the individual's age is the

variable creating the most doubts: the underlying preferences of individuals, reliably dissimilar among age groups, may strongly determine the speed of ceasing the labour contract, and thus the shape of the baseline hazard function. As mentioned in Section 1.5.1, the stratification approach add more flexibility by allowing people to experience different baseline hazards instead of being one the multiplicative version of the other[32]. Various post-estimation diagnoses have been initially performed on the baseline model 4 to inspect our concern. According to the PH assumption, the effects of covariates "do not change with time except in ways that you have already parameterized" (Cleves et al. (2010)). Therefore, by including time-dependent covariates in the model, i.e. the interactions between each predictor and a time component, we can easily verify whether or not interactions are different from zero. When significant predictors appear, a violation of the proportional assumption for that specific covariate arises. Among our age dummies, defined in Model 4 according to age-distribution, only the interaction between the first group (individuals aged 18 to 45) and time is found to be significant at 5% confidence level. An alternative way of testing the PH assumption is through the Schoenfeld and scaled Schoenfeld residuals: basically, when the PH assumption for a specific covariates holds, the Schoenfeld residual must not be related to survival times[33]. Both the proportionality test of predictors (p-values greater than 0.05) and the graphical check (Figure A3 in Appendix shows almost perfect horizontal lines) reject the violation of the PH assumption. Based upon these tests, unable to provide a clear and univocal answer to our concern, we get the conclusions by comparing the baseline Cox PH model with its stratified version. Table 6 compares the results with and without the stratification approach. As it is clear from column 2, there are virtually no differences when the hazards are constrained to be multiplicative replicas of each other (baseline esteems) or they are allowed to change freely among different age groups.

Similar concerns on the equality of baseline hazards also arise with respect to the type of cardiovascular shock occurred: the level of impairments derived from cerebrovascular diseases, namely strokes, can be potentially stronger and more severe than cardiovascular ones. According to this, the way how the two groups of blue-collar workers behave and thus, the speed of exit from that labour contract, can be very different. Table 6 column 3 shows the stratification approach applied on the type of CVD shock. As additional check, column 4 reports the results when both age and disease groups are considered. No changes appear, thus increasing the internal validity of the baseline results.

[32]The baseline Cox PH model allows people to differ in their covariates' values.

[33]The test proceeds as follows: the first step is to retrieve the residuals from the baseline estimation, and then, by fitting a smoothed function of time to them, the test will check whether a significant relationship turns out.

Table 6: Effect of sick leave duration on job-interruption - *Stratification*

		Stratification		
	baseline	*by age-groups*	*by CVD type*	*by age-group and CVD type*
sick_leave	1.016***	1.016***	1.016***	1.016***
	(0.004)	(0.004)	(0.004)	(0.004)
Past Health Characteristics	YES	YES	YES	YES
Current Job Characteristics	YES	YES	YES	YES
Past Job Characteristics	YES	YES	YES	YES
N.Obs.	1354	1354	1354	1354
Log pseudolikelihood	-6668.8	-5279.4	-6096.6	-4715.9
Wald chi2	395.2	182.3	393.52	169.24

Source: WHIP&Health. Notes: The table compares the *hazard ratios* of the last model in Table 3 (Model 4) with its stratified version: the stratification has been performed according to the age distribution of people (18-46/ 47-51/ 52-55 /56+). Robust standard errors have been considered. * p<0.1, ** p<0.05, ***p<0.01

We end this section by offering an overview of both the advantages and disadvantages of parametric duration models. As briefly mentioned before, while the Cox model can be performed without making any assumption about the shape of the baseline hazard function ($h_0(t)$), when the functional form is known, an efficiency gain can be obtained by its parameterization. Table A5 in Appendix makes clear how the differences between parametric and semi-parametric estimates are not remarkable *per se*: the magnitudes are similar. However, as an undeniable advantage, the former group of models permits to consider a random component - the frailty component (α_i) - allowing people in the population to differ due to unobserved factors[34]. In particular, when $\alpha_i < 1$ the hazard decreases, i.e. that individual is less risky than others; on the contrary, if $\alpha_i > 1$ he/she is characterised by a higher risk to frail. The frailty component, not exploited by Cox models[35], has a multiplicative effect on hazard and it is assumed to follow a specific distribution with mean equal to 1 and variance θ. Table 7 compares two types of distributions, the gamma and the inverse-Gaussian: the choice implies a different interpretation of how the relative hazard changes with time. Rather than the magnitude and significance associated with sick leave, two main values are extremely important here. The first one is the measure of θ, i.e. the estimated variance of the frailty component, and the second is the p-value of the Likelihood Ratio Test. Under the null hypothesis of the LR test, the variance is equal to zero and the frailty component does not contribute to the model. According to it, if we are willing to accept that the individual's hazard moves as a Weibull distribution, then there is evidence pointing toward a heterogeneous population. The estimated coefficient for sick leave cannot be directly compared with the previous results. Although they are still hazard ratios, their interpretation when frailty is included is slightly different: 1.027 (or 2.7%) is the estimated hazard of a unitary increase of sick leave between two individuals sharing the same frailty.

[34]Shortly, the frailty model can be represented as follows:

$$h(t_i|\mathbf{X}_i, \alpha_i) = \alpha_i h(t_i|\mathbf{X}_i)$$

where α_i is the unobserved individual-specific effect.

[35]The shared frailty is the only option of Cox models. With shared frailty models clusters of subjects are assumed to share the same frailty. For example, subjects from the same family may be similar with respect to some unobserved genetic. (Kleinbaum et al. (2005)

Table 7: Effect of sick leave duration on job-interruption - *Parametric models with frailty*

	Model 4 (Weibull)	
	Gamma	*Inverse-Gaussian*
sick_leave	1.027***	1.028***
	(0.006)	(0.007)
N.Obs.	1354	1354
p	1.117	1.247
θ	0.724	2.884
LR test	0.000	0.000

Source: WHIP&Health. Notes: The table compares the *hazard ratios* of two types of parametric models with frailty component: column 1 considers a gamma frailty distribution while column 2 an inverse-Gaussian distribution. Under the null of the Likelihood Ratio test $\theta = 0$. No robust standard errors. * p<0.1, ** p<0.05, ***p<0.01

Overall, the choice between a parametric model with heterogeneity control and a semi-parametric model is not an obvious one (Hesselius (2007)): the former assumes independence between individual heterogeneity and covariates, likely to be violated. Instead, the latter has been shown by Lancaster (1990) to yield a bias toward zero when the unobserved heterogeneity is neglected by partial likelihood estimates. As stressed in the previous sections, Cox models are the safest choice especially when the modelled risky process - in this analysis, the time until job-interruption occurs - follows an ambiguous dynamic over time and thus, the shape of the hazard cannot be safely parametrized.

8 Conclusion

This paper analyses the role played by sick leave duration, under severe health conditions, on the risk of exit from a specific (and permanent) job. Although sick leave is an instrument designed by policymaker to prevent the potential income losses related to bad health, it is also a channel through which the employer can be negatively warned. 'The previous absence behaviour of workers can be seen as a signal for the employer or future employers of worker's health status and/or shirking tendency' (Hesselius (2007)). Thus, as extensively proved by the literature, as sick leave increases, the future risk of unemployment also increases. By assuming a slightly different perspective, the research looks at the labour contract which in place at the time of the reference CVD hospitalisation. As permanent employees, their working activity should be highly supported, both in terms of dismissal and intra-firm reallocation of tasks when health status prevent suitable performances. Upon this background, the effectiveness of the employment protection legislation (EPL), rather heterogeneous by firm dimension, is extremely relevant, especially when blue-collar workers are considered. From a broad analysis of EU-LFS data is evident how the type of job can influence the post-shock return-to-work (Figure A1 and A2 in Appendix). Overall, the way in which sick leave may affect the residual job-tenure is not clearly determined; opposite forces can play a role when permanent workers are studied.

Thanks to a novel administrative dataset, the aim of this paper has been addressed by

focussing on male blue-collar workers hit by a severe form of cardiovascular disease between 2003 and 2005. Through the specific sample selection, a sizeable part of the endogeneity issues arising from sick leave can be relaxed. Moreover, detailed information about the date of the reference CVD hospitalisation and the closing date of the labour contract make the continuous-time duration models as the most suitable in this context. In particular, the original version of the Cox proportional hazard model with time-invariant covariates (Cox, 1972), has been performed. The baseline results show a negative relationship between sick leave duration and the subsequent job-interruption: an additional week of absence increases the instantaneous risk of exit from the labour contract of about 1.6%. It is worth mentioning that both the magnitude and significance are insensitive to controls (models 1 to 4), suggesting how the 'the time needed to return-to-work' does not reflect additional confounding factors. Besides, we further investigate our baseline findings by including two main interaction terms. Heterogeneities among different age groups and different dismissal incentives according to the firm dimension may potentially affect the role of extended periods at home. Surprisingly, the effect of an extended period at home after the experience of an acute CVD shock has similar (negative) consequences along with all age groups and the firm's dimension. Regardless of the number of weeks at home, bigger firms offer greater opportunities to continue a regular working activity: being employed in small companies increases the instantaneous risk of job-interruption of about 43%. A variety of sensitivity checks and alternative survival approaches increase the confidence of our findings.

Overall, the paper offers a worrisome picture of the limited working opportunities unhealthy blue-collars face after an acute health shock. Surprisingly, being permanently employed is not enough to cope with these growing difficulties: in other words, the guarantees offered by the employment protection legislation (EPL) seem insufficient to allow blue-collar workers a safe continuation of their jobs. In a context where the rate of incidence of CVD diseases is increasing while the rate of mortality is decreasing (EHN, 2017), those labour markets unable to facilitate these types of workers will be strongly under pressure. Therefore, more targeted policies aimed to help specific categories of workers, together with a specific attention on the type of disease, should also be implemented. Our findings are in line with that part of the literature pointing the lack of workplace arrangements as one of the leading cause of job-interruption (Hill et al. (2016), Anand et al. (2017)). The employer's trade-off between the costs of workplace adaptation, legal constraints and the potential costs of dismissal, undoubtedly determine the likelihood as well as the speed of a job-interruption.

References

[1] Anand, P., Sevak, P. (2017). *The role of workplace accommodations in the employment of people with disabilities.* IZA Journal of Labor Policy, 6.

[2] Arai M., Thoursie P.S. (2005) *Incentives and selection in cyclical absenteeism.* Labour Economics 12(2), 269-280

[3] Askildsen J.E., Bratberg E., Nilsen I.A. (2005) *Unemployment, labor force composition and sickness absence: a panel study.* Health Economics, 14 (2005), pp. 1087-1101

[4] Barmby T., Marco A., Ercolani G., Treble J.G. (2002) *ickness absence: an international comparison.* Economic Journal, 112 (480) (2002), pp. F315-F331

[5] Biscardo C.A., Bucciol A., Pertile P. (2019) *Job sick leave: Detecting opportunistic behavior.* Health Economics; 28(3):373-386.

[6] Boeri T.Jimeno J. (2005) *The effects of employment protection: learning from variable enforcement.* European Economic Review , 49, 2057-77

[7] Burkhauser R.V., Butler J.S., Kim Y.W. (1995). *The importance of job accommodation on the job duration of workers with disabilities: a hazard model approach.* Labour Economics, 2 (1995), pp. 109-130

[8] Burkhauser, R. V., Butler, J. S., Kim, Y. W., Weathers, R. R. (1999). *The importance of accommodation on the timing of disability insurance applications: Results from the survey of disability and work and the health and retirement study.* Journal of Human Resources, 589-611.

[9] Campolieti, M. (2005). *How accommodations affect the duration of post?injury employment spells.* Journal of Labor Research, 26, 485-499.

[10] Cleves M., Gutierrez R.G., Gould W., Marchenko Y. V. (2010) *An introduction to survival analysis using stata* Stata Press book, third edition

[11] Contini, B. (2019), Dai "lavori usa e getta" al Jobs Act. Aracne editrice, Roma.

[12] Cox D.R. (1972) *Regression models and life-tables* J. R. Stat. Soc. Series B (Methodological), 34 (2) (1972), pp. 187-220

[13] De Paola M. , Pupo V. , Scoppa V. (2014) *Absenteeism in the Italian public sector: the effects of changes in sick leave policy.* Journal of Labor Economics

[14] Hesselius P (2007) *Does sickness absence increase the risk of unemployment?.* Journal of Socio-Economics 36: 288-310.

[15] Hill, M., Maestas, N., Mullen, K. J. (2016). *Employer accommodation and labor supply of disabled workers*. Labour Economics, 41, 291-303.

[16] Ichino A., Riphahn R.T. (2005) *The effect of employment protection on worker effort. A comparison of absenteeism during and after probation*. Journal of European Economic Association, 3 (1) (2005), pp. 120-143

[17] Jenkins S. P. (2005) *Survival Analysis* Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.176.7572&rep=rep1&type=pdf

[18] Johansson P., Palme M. (2002) *Assessing the effect of public policy on worker absenteeism*. Journal of Human Resources, 37 (2) (2002), pp. 381-409

[19] Kleinbaum D. G., Klein M. (2005). *Survival Analysis. A self-Learning Text*. Springer Book, second edition

[20] Lancaster T. (1990) *The econometric analysis of transition data* Cambridge University Press, Cambridge

[21] Leombruni, R., (2011). *Livello di protezione del contratto e assenteismo* Atti del Convegno Absenteeism in the Italian Public and Private Sector: The Effects of Changes in Sick Leave Compensation, 159-168.

[22] Markussen S., Roed K., Rogeberg O., Gaure S. (2011) *The anatomy of absenteeism*. Journal of Health Economics, 30 (2) (2011), pp. 277-292

[23] Markussen S. (2012) *The individual cost of sick leave*. J. Popul. Econ., 25 (4) (2012), pp. 1287-1306

[24] Neumark D, Bradley CJ, Henry M, Dahman B. (2015). *Work continuation while treated for breast cancer: The role of workplace accommodations*. ILR Review, 68, 916- 954.

[25] OECD (1999). *The Employment Outlook*, Chapter 5

[26] Pfeifer C. (2013) *Cyclical absenteeism among private sector, public sector and self-employed workers*. Health Economics, 22 (3) (2013), pp. 366-370

[27] Puhani P.A., Sonderhof K. (2010) *The effects of a sick pay reform on absence and on health-related outcomes*. Journal of Health Economics, 29 (2) (2010), pp. 285-302

[28] Schön, M. (2015) *Unemployment, sick leave and health.*

[29] Scoppa V., Vuri D. (2014) *Absenteeism, unemployment and employment protection legislation: Evidence from Italy*. IZA J Labor Econ (2014) 3: 3. https://doi.org/10.1186/2193-8997-3-3

[30] Spasova S., Bouget D. and Vanhercke B. (2016). *Sick pay and sickness benefit schemes in the European Union*. Background report for the Social Protection Committee?s In-depth Review on

sickness benefits (17 October 2016), European Social Policy Network (ESPN), Brussels, European Commission.

[31] Wilkins E, Wilson L, Wickramasinghe K, Bhatnagar P, Leal J, Luengo-Fernandez R, Burns R, Rayner M, Townsend N (2017). *European Cardiovascular Disease Statistics 2017* European Heart Network, Brussels.

[32] Ziebarth, N. R. (2013) *Long-term Absenteeism and Moral Hazard: Evidence from a Natural Experiment*. Labour Economics 24: 277-292.

Appendices

Table A1: Frequencies by diagnoses

ICD 9-CM diagnostic category	**Num.**	%
Ischemic Heart Disease	**(1029)**	**(76.00%)**
Acute myocardial infarction (410)	533	39.36
Other acute and subacute forms of ischemic heart disease (411)	202	14.92
Old myocardial infarction (412)	10	0.74
Angina pectoris (413)	149	11.00
Other forms of chronic ischemic heart disease (414)	135	9.97
Cerebrovascular Disease	**(325)**	**(24.00%)**
Subarachnoid hemorrhage (430)	24	1.77
Intracerebral hemorrhage (431)	29	2.14
Other and unspecified intracranial hemorrhage (432)	8	0.59
Occlusion and stenosis of precerebral arteries (433)	34	2.51
Transient cerebral ischemia (434)	114	8.42
Other and ill-defined, cerebrovascular disease (436)	39	2.88
Late effects of cerebrovascular disease (437)	77	5.69
Total number of admissions	1354	100.00

Source:WHIP&Health

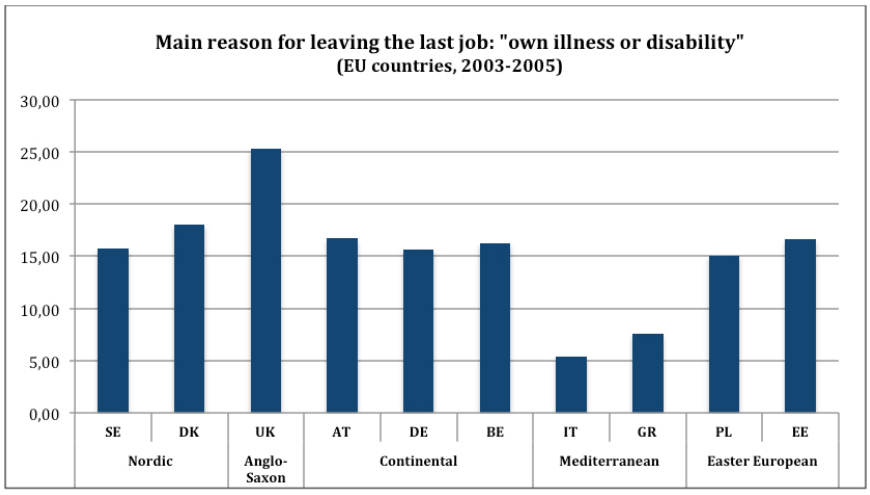

Figure A1: EU comparison of male blue-collar workers reporting "own illness or disability"
Source: EU-LFS. Notes: Percentages of male blue-collar workers reporting "own illness or disability" as the main reason for leaving the last job. EU countries are grouped according to their homogeneity in cultural attitudes, social security environment, labour and welfare institutions

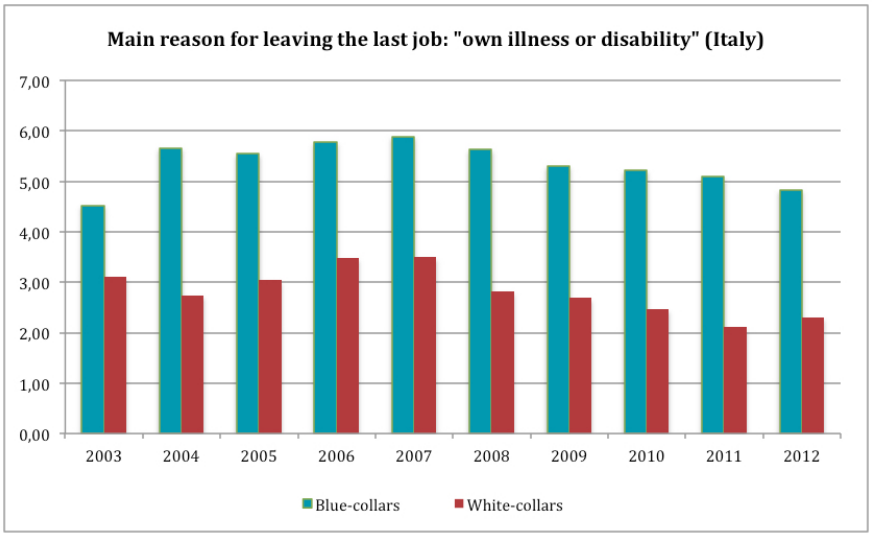

Figure A2: Differences among those answering "own illness or disability" over time and type of worker
Source: EU-LFS. Notes: Differences in the percentage of male blue and white collar workers reporting "own illness or disability" as the main reason for leaving the last job. EU countries are grouped according to their homogeneity in cultural attitudes, social security environment, labour and welfare institutions

Table A2: Cox proportional hazard model - *baseline*

	Model 1	Model 2	Model 3	Model 4
sick_leave	1.017***	1.018***	1.015***	1.016***
	(0.004)	(0.004)	(0.004)	(0.004)
age1846	0.336***	0.337***	0.335***	0.331***
	(0.033)	(0.033)	(0.033)	(0.033)
age4751	0.368***	0.369***	0.389***	0.390***
	(0.035)	(0.035)	(0.037)	(0.038)
age5255	0.664***	0.665***	0.674***	0.670***
	(0.054)	(0.054)	(0.058)	(0.060)
abirth_centre	1.000	0.997	0.964	0.976
	(0.102)	(0.102)	(0.158)	(0.160)
abirth_south	1.074	1.070	1.081	1.081
	(0.085)	(0.085)	(0.117)	(0.121)
abirth_islands	1.368**	1.364**	1.325**	1.281**
	(0.138)	(0.138)	(0.158)	(0.152)
abirth_abroad	0.883	0.876	0.715	0.703
	(0.344)	(0.343)	(0.322)	(0.306)
country_underdev	1.212	1.215	1.321	1.281
	(0.488)	(0.490)	(0.604)	(0.563)
days_other_$\bar{t}$	1.010***	1.010***	1.010***	1.010***
	(0.002)	(0.001)	(0.002)	(0.002)
days_cvd_$\bar{t}$	0.996	0.996	0.998	0.997
	(0.007)	(0.007)	(0.007)	(0.007)
sick_leave_cum		0.999	1.001	1.000
		(0.001)	(0.001)	(0.001)
inv_benefit_cum		1.004	1.003	1.002
		(0.017)	(0.016)	(0.016)
days_other_cum		1.002	1.004	1.003
		(0.005)	(0.004)	(0.005)
days_cvd_cum		1.004	0.997	0.991
		(0.02)	(0.021)	(0.023)
lab_income(*log*)			0.746**	0.758**
			(0.079)	(0.083)
m_seniority			0.999***	0.999
			(0.0004)	(0.001)
part_time			0.586***	0.663**
			(0.099)	(0.116)
s_secondary			0.869	0.847
			(0.115)	(0.111)
s_tertiary			0.863	0.892
			(0.124)	(0.127)
arwork_centre			1.002	0.975
			(0.137)	(0.135)
arwork_south			0.975	1.012
			(0.098)	(0.103)
firm_015			1.439***	1.431***
			(0.112)	(0.112)
work_active_cum				0.927***
				(0.024)
nemployee_cum				1.056**
				(0.021)
ever_selfempl				1.149**
				(0.178)
ever_atypical				1.354
				(0.280)
nunempl_cum				0.929***
				(0.027)
ever_cig				1.213**
				(0.089)
N.Obs.	1354	1354	1354	1354
Log pseudolikelihood	-6722.4	-6722.1	-6683.8	-6668.8
Wald chi2	283.0	286.3	369.3	395.2

Source: WHIP&Health. Note: Extended version of Table 3: comparison of the *hazard ratios* of four Cox PH models where different covariates have been included. The variable "age_5664" is the reference age group; "abirth_north" is the reference group for area of birth; "s_primary" is the reference sector of activity; "awork_north" is the reference area of work. Robust standard errors have been considered. * p<0.1, ** p<0.05, ***p<0.01

Table A3: Cox proportional hazard model - *Interactions*

	No interactions (Model 4)	**Interaction1**	**Interaction 2**
sick_leave	1.016***	1.021**	1.016***
	(0.004)	(0.008)	(0.005)
age1846	0.331***	0.368***	0.331***
	(0.033)	(0.060)	(0.033)
age4751	0.390***	0.500***	0.390***
	(0.038)	(0.079)	(0.038)
age5255	0.670***	0.667**	0.690***
	(0.060)	(0.102)	(0.060)
sickleave_1846	-	0.991	-
	-	(0.012)	-
sickleave_4751	-	0.980*	-
	-	(0.012)	-
sickleave_5255	-	1.002	-
	-	(0.011)	-
firm_015	1.431***	1.431***	1.429***
	(0.112)	(0.112)	(0.177)
sickleave_firm015	-	-	1.000
	-	-	(0.009)
abirth_centre	0.976	0.987	0.975
	(0.160)	(0.161)	(0.160)
abirth_south	1.081	1.080	1.081
	(0.121)	(0.122)	(0.121)
abirth_islands	1.281**	1.130**	1.280**
	(0.152)	(0.154)	(0.152)
abirth_abroad	0.703	0.743	0.703
	(0.306)	(0.325)	(0.307)
country_underdev	1.281	1.212	1.281
	(0.563)	(0.537)	(0.565)
days_other_t	1.010***	1.010***	1.010***
	(0.002)	(0.002)	(0.002)
days_cvd_$\bar{t}$	0.997	0.997	0.997
	(0.007)	(0.007)	(0.007)
sick_leave_cum	1.000	1.000	1.000
	(0.001)	(0.001)	(0.001)
inv_benefit_cum	1.002	1.003	1.002
	(0.016)	(0.017)	(0.016)
days_other_cum	1.003	1.003	1.003
	(0.005)	(0.005)	(0.005)
days_cvd_cum	0.991	0.990	0.992
	(0.023)	(0.022)	(0.023)
lab_income(log)	0.758**	0.754**	0.758**
	(0.083)	(0.083)	(0.083)
m_seniority	0.999	0.999	0.999*
	(0.001)	(0.001)	(0.001)
part_time	0.663**	0.660**	0.663**
	(0.116)	(0.117)	(0.115)
s_secondary	0.847	0.850	0.847
	(0.111)	(0.111)	(0.127)
s_tertiary	0.892	0.904	0.892
	(0.127)	(0.128)	(0.127)
arwork_centre	0.975	0.965	0.975
	(0.135)	(0.133)	(0.135)
arwork_south	1.012	1.007	1.012
	(0.103)	(0.103)	(1.103)
work_active_cum	0.927***	0.928***	0.927***
	(0.024)	(0.024)	(0.024)
nemployee_cum	1.056**	1.056**	1.056**
	(0.021)	(0.021)	(0.021)
ever_selfempl	1.149**	1.137	1.149
	(0.178)	(0.177)	(0.177)
ever_atypical	1.354	1.322	1.354
	(0.280)	(0.277)	(0.280)
nunempl_cum	0.929***	0.930**	0.929**
	(0.027)	(0.027)	(0.027)
ever_cig	1.213**	1.214**	1.213**
	(0.089)	(0.090)	(0.089)
N.Obs.	1354	1354	1354
Log pseudolikelihood	-6668.8	-6666.4	-6668.8
Wald chi2	395.2	399.7	395.8

Source: WHIP&Health. Note: Extended version of Table 4: comparison of the *hazard ratios* of the baseline Cox PH model and two alternative specifications where interaction terms are included. The variable "age_5664" is the reference age group; "abirth_north" is the reference group for area of birth; "s_primary" is the reference sector of activity; "awork_north" is the reference area of work. Robust standard errors have been considered. * p<0.1, ** p<0.05, ***p<0.01

Table A4: Effect of sick leave duration on job-interruption - *Sensitivity Check*

	Baseline (Model 4)	No hospitalisations at time T
sick_leave	1.016***	1.010
	(0.004)	(0.007)
Past Health Characteristics	YES	YES
Current Job Characteristics	YES	YES
Past Job Characteristics	YES	YES
N.Obs.	1354	677
Log pseudolikelihood	-6668.8	-2884.34
Wald chi2	395.15	241.21

Source: WHIP&Health. Notes: The table compares the *hazard ratios* of the last model in Table 3 (Model 4) with an additional specification of Cox PH model: we exclude those who experienced additional hospitalisations for CVD shock and other types of diseases in the same year as the referenced CVD hospitalisation. Robust standard errors have been considered. * p<0.1, ** p<0.05, ***p<0.01

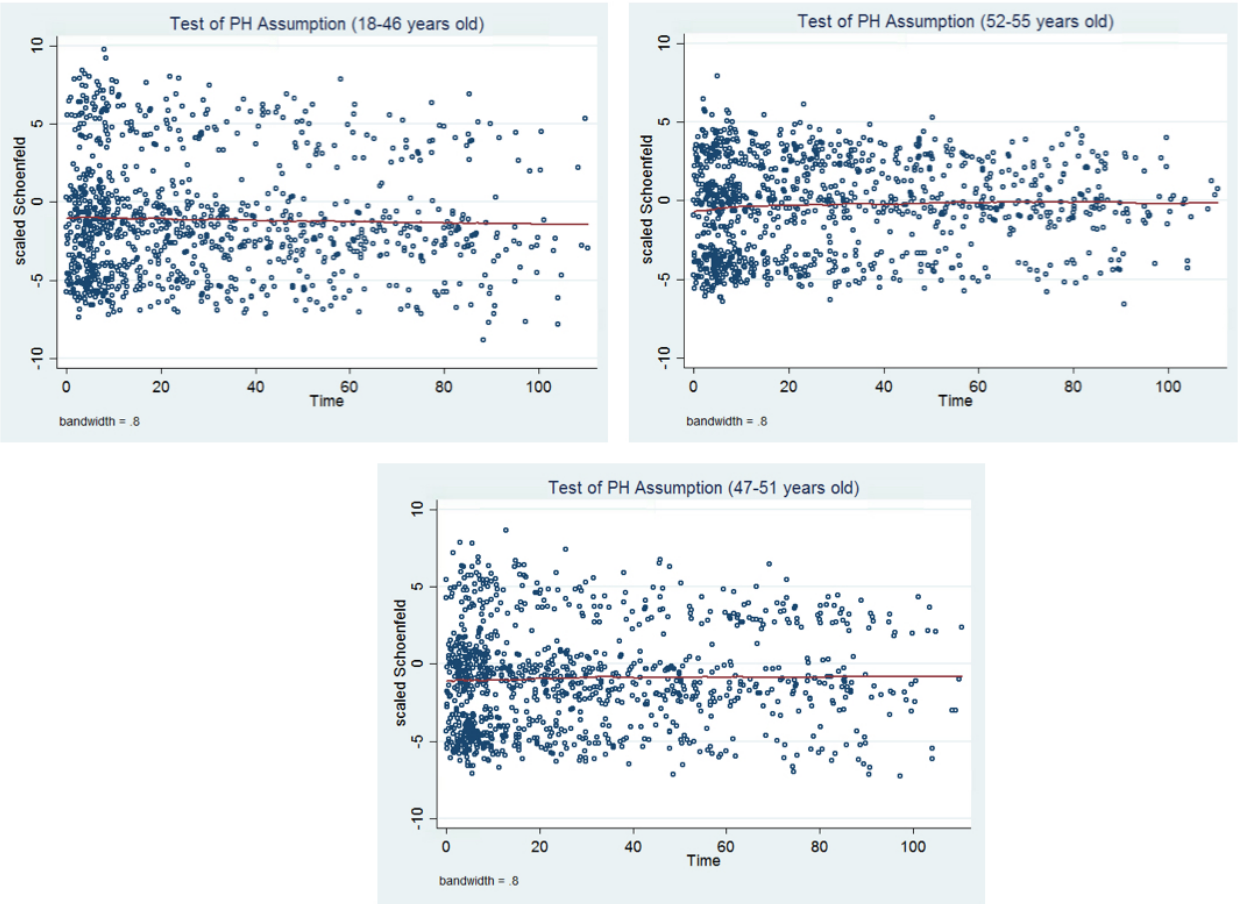

Figure A3: Test of PH assumption by age groups - Schoenfeld residuals
*Source: WHIP&Health.***Notes:** The idea of this test is to retrieve the residuals, fit a smooth function of time to them, and then test whether there is a relationship (Cleves et al. 2010). Model 4 in Table 3 is our reference estimation. The baseline age group are those between 56 and 64 years old.

Table A5: *Parametric Estimates - no frailty*

	Baseline *Cox Model*	Exponential	Weibull	Gompertz
sick_leave	1.016***	1.016***	1.015***	1.015***
	(0.004)	(0.004)	(0.004)	(0.004)
Past Health Characteristics	YES	YES	YES	YES
Current Job Characteristics	YES	YES	YES	YES
Past Job Characteristics	YES	YES	YES	YES
N.Obs.	1354	1354	1354	1354
Log pseudolikelihood	-6668,8	-2158.1	-2141.2	-2140.7
LR(chi2)	381.22	477.56	398.28	396.64

Source: WHIP&Health. Note: The table compares the *hazard ratios* of our baseline semi-parametric duration model, i.e. the Cox Model (Model 4, Table 3), with its parametric version. Survival times have been modelled according to three different distributions: *exponential*, *Weibull* and *Gompertz*. According to the distributions, the hazard function assumes the following forms respectively: $h(t) = \lambda$, $h(t) = \lambda p t^{p-1}$ and $h(t) = exp(\gamma t)$. * p<0.1, ** p<0.05, ***p<0.01