The economics of health and safety: An interdisciplinary review of the theory and policy

Konstantinos, Pouliakas and Ioannis, Theodossiou

University of Aberdeen Business School, IZA

28 October 2010
THE ECONOMICS OF HEALTH AND SAFETY AT WORK: AN INTERDISCIPLINARY REVIEW OF THE THEORY AND POLICY

Konstantinos Pouliakas¹
CELMR, University of Aberdeen Business School & IZA

Ioannis Theodossiou²
CELMR, University of Aberdeen Business School

Abstract
This paper engages in an interdisciplinary survey of the current state of knowledge related to the theory, determinants and consequences of occupational safety and health (OSH). It first describes the fundamental theoretical construct of compensating wage differentials, which is used by economists to understand the optimal provision of OSH in a perfectly competitive labour market. The plethora of incentives faced by workers and firms in job and insurance markets that determine the ultimate level of OSH are discussed in detail. The extensive empirical evidence from the hedonic wage and stated choice approaches used to assess the value of OSH is reviewed. The causes of inefficiency and inequity in the market for OSH, such as externalities, moral hazard in compensation insurance, systematic biases in individual risk perception/well-being and labour market segregation are subsequently examined. The implications of government intervention and regulation for tackling the aforementioned inefficiencies in OSH are then considered. Finally, the survey identifies areas of future research interests and suggests indicators and priorities for policy initiatives that can improve the health and safety of workers in modern job markets.

Keywords: health, safety, accidents, illnesses, absenteeism, regulation

JEL-Codes: J17, J28, J81, K32

¹ Research Fellow, IZA and Centre for European Labour Market Research (CELMR), University of Aberdeen Business School, Edward Wright Building, Dunbar Street, Old Aberdeen AB24 3QY, UK – Tel. ++44 01224 272172, email: k.pouliakas@abdn.ac.uk.
² Corresponding author: Professor of Economics, Centre for European Labour Market Research (CELMR), University of Aberdeen Business School, Edward Wright Building, Dunbar Street, Old Aberdeen AB24 3QY, UK – Tel. ++44 01224 272181, email: theod@abdn.ac.uk.
1. Introduction

"All too often lives are shattered unnecessarily of poor working conditions and inadequate safety systems. Let me encourage everyone to join the ILO in promoting safety and health at work. It is not only sound economic policy; it is a basic human right".

Mr. Kofi Annan, former Secretary General of the United Nations

The purpose of this survey is to review the current state of knowledge and issues related to theory and empirical evidence of the market for occupational safety and health (OSH). The increasing competition prompted by globalisation, the predominance of service-oriented industries, the rising job insecurity associated with labour market flexibility and demographic developments in the composition of the workforce (e.g. ageing, feminization), pose important challenges for the health and safety of workers in modern economies. In addition, many governments have recently paid greater attention to the need to tackle the non-trivial costs to both individual and societal welfare that the lack of OSH entails, as part of their overall strive to overhaul insolvent social security regimes.

An indication of the considerable costs associated with the lack of provision of OSH is given by estimates of large international bodies such as the World Health Organization (WHO) and the World Bank. They attribute about 3% of lost life years to the factor ‘work’ (Kreis and Bodeker, 2004). Furthermore, social insurance expenditure on OSH (e.g. statutory sick pay, disability allowances, industrial injuries disablement and incapacity benefits) accounts for approximately 2-3% of Gross Domestic Product (GDP) in most advanced economies, exceeding by far what is typically spent on unemployment benefits (Adema and Ladaigue, 2009). International Labour Organization (ILO) estimates also show that work-related diseases and accidents account for economic losses as high as 4% of world-wide GDP (ILO, 2003). Around 4 million accidents at work resulting in more than 3 days of absence occurred in the EU-15 in 2005 (European Commission, 2009, p. 19). This corresponds to an incidence rate of 3100 non-fatal and 3.5 fatal accidents per 100,000 workers. Furthermore, for each worker in the EU-15 an average of 1.3 working days is lost each year due to an accident at work and 2.1 days are lost because of other work-related health problems (European Communities, 2004). 2-4% of contracted work hours are also estimated to be lost due to sickness absence (Adema and Bonato, 2007). Some calculations suggest that the socio-economic costs of (sickness) absence in advanced Western economies account for between 2-3% of their total GDP (EUROFOUND, 1997), that is a typical year’s growth. Similarly, the estimated direct and indirect costs of work-related injuries and illnesses in the US are approximately $170 billion annually (CSTE, 2005; CDC, 2007). All of the above figures neglect other major non-quantifiable costs, such as private insurance and health care outlays that affected individuals face, the indirect costs that companies incur (e.g. training inexperienced replacement workers, administrative expenses, production bottlenecks, low employee morale), the impact on families and communities and the inefficiency of having a large proportion of a potentially active workforce disabled, idle or prematurely retired.

The equilibrium level of OSH in economies is determined by the interplay of incentives faced by workers and firms in labour and insurance markets. This level is influenced by government regulation that is implemented in order to tackle any market inefficiencies that may arise. To the extent that the above costs reflect market inefficiencies, the need for up-to-date information on OSH is of critical importance in order to identify areas of required action and to set priorities for policy initiatives on improving health and safety at work. According to recent ILO (2008) estimates, the global number of work-related fatal and non-fatal accidents and diseases has been relatively stable during the past 10 years, despite the significant improvement in OSH attained in many developed economies. This is mainly due to the globalization process and rapid industrialization that have led to a deterioration of OSH performance in relatively poor countries which are unable to maintain effective OSH standards. Hence the need to focus on health and safety is paramount, given that “the traditional hazard and risk prevention and control tools are still effective but need to be completed by strategies designed to address the consequences of a continuous adaptation to a rapidly changing world of work” (ibid, 2008, p. vii).

The purpose of this survey is, thus, to review and augment current knowledge on the economics of occupational health and safety (Viscusi, 1993; Brown and Sessions, 1996; Shapiro, 1999; Ruser and Butler, 2009) by utilizing a comprehensive multidisciplinary approach. By synthesizing the available economic literature with evidence from
disciplines such as occupational medicine, epidemiology, psychology and sociology, the survey provides a holistic overview of the market for OSH and of its implications for individual workers, companies, economies and societies. This is important, since OSH constitutes an area of lively discourse across disciplines, but complementarities in the methodological or empirical findings have not yet been brought together (Weil, 2001). The review therefore employs an integrative analysis in order to provide a useful guide to researchers and policymakers, who may be bewildered by the voluminous but disparate amount of publicly available information on the issue of OSH.3

The structure of the paper is as follows. Section 2 outlines the fundamental theoretical construct used by economists for analyzing OSH equilibrium in a perfectly competitive market. The determinants of the plethora of incentives faced by workers and firms in job and insurance markets and the extensive empirical evidence from the hedonic wage and stated choice approaches used to assess the value of OSH are discussed in Section 3. Section 4 focuses on the causes of inefficiency and inequity in the market for OSH, such as externalities, moral hazard in compensation insurance, systematic biases in individual risk perception/well-being and labour market segregation. Section 5 examines the implications of government intervention for tackling the aforementioned inefficiencies. Finally, Section 6 concludes by reviewing the major findings that can inform public policy on OSH and by identifying areas of future concern.

2. The Market for OSH: Theoretical Underpinnings

The level of OSH in an economy is determined by the interplay of incentives faced by workers and firms in job and insurance markets and is moderated by the regulatory activities of government. Economists have typically analyzed the safety decisions of both employees and firms in a perfectly competitive market as the outcome of a complex set of factors reflecting various costs and benefits of investing in OSH (Viscusi, 1993; Shapiro, 1999; Ruser and Butler, 2009). On the one hand, workers are assumed to be rational and fully-informed agents who are likely to demand a “wage premium” as compensation for OSH risks. Their safety activities on the job and their degree of absenteeism once ill or injured is also believed to be affected by their exposure to social insurance and compensation benefit schemes. On the other hand, employers are assumed to face a trade-off between the cost of investing in health and safety precautions and the expected benefits of such actions, most notably lower wages offered to employees as compensation for job disamenities or risk. The equilibrium level of OSH in the economy is therefore the outcome of the interaction between the labour demand and labour supply decisions of firms and workers, respectively.

2.1 The Market for OSH: Heterogeneity in market opportunities offered by employers

In the spirit of Henderson (1983), a convenient formalisation of the safety decisions faced by employers in market economies is shown in Figure 1. In the absence of regulation, a sufficient condition for profit-maximising firms selecting the optimal safety level, $S_f^*$ (or, equivalently, the optimal job risk level, $R_f^*$), is that their OSH decisions should equalize their OSH costs and OSH benefits at the margin. The ante-factum outlays associated with improvement in workplace conditions (e.g. preventative practices, OSH training) are likely to rise as the level of health and safety increases. This is reflected in the convexity of the total costs (TC) curve in Figure 1. The post-factum benefits (or, equivalently, reduced damage costs) are associated with having to pay lower (‘compensating’) wages and workers’ compensation due to the reduction of workplace injuries and illnesses, a lower level of sickness absence and sick pay, reduced insurance premiums, and the evasion of other non-trivial costs to the firm (disruption in production, replacement of specifically-trained workers, low worker morale, bad reputation etc.). These are likely to fall at a diminishing rate as OSH levels increase, as shown by the downward-sloping total benefits (TB) curve, since the severity of incidents at higher levels of safety are subdued. As illustrated in Figure 1, only at $S_f^*$ ($R_f^*$) is the cost of extra safety equalised with the value of the benefits that an inframarginal unit of job safety entails for a given firm.

The above reasoning implies that in order to maintain constant profits, firms must offer lower wages ($w$) to offset any marginal addition in the cost of OSH. The wage offer of a given firm is therefore an increasing function of job risk, as is depicted by the concave isoprofit curve (II) in Figure 2. Furthermore, the optimal wage-risk pairs ($R_f, w$) that can maintain the same level of profits are likely to vary across different firms, given that they face a diverse set of...
technologies and job hazards. Working environments differ with respect to the exposure of employees to various inputs of the production process, most notably physical agents (e.g. noise, vibration, radiation, room temperature), chemical agents (e.g. asbestos, lead, benzene, pesticides), poor ergonomics (e.g. inconvenient work postures, repetitive movements, lifting of heavy materials), adverse working times (long and irregular workdays, shifts, night work) and workplace violence (bullying, harassment, discrimination). For instance, smaller-sized firms and ‘blue-collar’ industries with a predisposition to the use of heavy metallic instruments, like engineering or shipyards, exhibit a greater degree of noise and vibration at work compared to other firms (Mery, 1973; EU-OSHA, 2000; Pyykko et al., 2007). The occupational medicine literature also highlights the negative health and safety effects due to contact with chemical substances in various trades (White and Proctor, 1997; Wong and Trent, 1999). Notable examples include the use of pesticides by agricultural workers (Hanke and Jurewicz, 2004) and the possibility of neural disorders suffered by workers in manufacturing jobs requiring regular contact with solvents, paints and various organic substances (Grasso et al., 1984; Wang and Chen, 1993; Kaukiainen et al., 2008). Carcinogen substances (e.g. acrylamide, asbestos, benzene etc.) are also particularly prevalent in industrial sectors (Collins et al., 1989; Marsh et al., 1999; Swaen et al., 2007). Moreover, workers employed in the manufacturing and health and social work sectors are frequently found to be affected by musculoskeletal disorders (e.g. repetitive strain injury (RSI), carpal tunnel syndrome and tenosynovitis), given that they are often required to engage in repetitive movements (EU-OSHA, 2000, p. 38-40). Finally, individuals working in service-sector occupations that rely on interacting with the public (e.g. health and social work, hotels and restaurants, education and public administration) are at higher risk of suffering from physical or mental violence, bullying and (sexual/racial) harassment at the workplace (Cassitto et al., 2003).

It is therefore clear that firms differ in their ability to trade-off the costs of providing a safer working environment with the expected benefits (particularly, a lower wage bill). This is illustrated by the heterogeneity in the wage offer curves II and JJ of two hypothetical firms in Figure 2.

2.2 The Market for OSH: Heterogeneity in preferences of workers for job risk

On the supply side of the market it is assumed that rational workers, who have perfect information and exhibit a high degree of labour market mobility, are likely to demand a wage premium as compensation for their willingness to endure danger on the job that is not covered by their ex post accident or sickness insurance. In order to maintain expected utility constant, a given worker will therefore demand alternative levels of wage compensation for varying degrees of job risk, as summarized by the $U_1U_1$ curve in Figure 2. However, workers exhibit differential tolerances to risk for a variety of reasons related to economic and demographic circumstances or simply differences in tastes (Ruser and Butler, 2009, p. 305). This heterogeneity in workers’ willingness to be employed in disagreeable job environments can therefore be depicted in Figure 2 by the dissimilar expected utility loci of two hypothetical workers, $U_1U_1$ and $U_2U_2$.

2.3 The Market for OSH: Compensating wage differentials for job risk and other disamenities

Superimposing the labour supply choices of individuals on the wage offer curves of firms results in what is known as “the fundamental long-run market equilibrium construct” of labour economics (Rosen, 1986), namely the theory of compensating wage differentials (CWDs). Rooted in the insightful work of Adam Smith’s Wealth of Nations (1776), the theory predicts that market forces will ensure the payment of wage premiums by firms which are characterised by inferior working conditions, as a means of recruiting and retaining valuable labour. It is further postulated that in a perfectly competitive labour market a positive equilibrium wage-risk relationship (shown by line EE) should arise due to the “matching” of the preferences of workers and firms. This is shown by the points of tangency between the expected utility loci and the market wage opportunity curves, ($r_1$, $w_1$) and ($r_2$, $w_2$), in Figure 2. Risk-averse workers are thus expected to take up jobs in firms which find it optimal to provide a safer work environment, whereas less risk-averse workers are more willing to be employed by firms which face a dearer marginal cost of safety provision. Such an assortative matching procedure predicts that jobs characterised by a higher degree of job risk (or other disamenities) should, in equilibrium, offer compensating wage rents, ceteris paribus. Therefore, market forces ensure that CWDs raise the cost of non-OSH provision to firms. This provides them with an inherent incentive to provide an adequate level of health and safety to their workforce.

As discussed in the previous section, at a given wage rate the optimal level of safety selected by firms, \( S_f^* \), occurs at the point where their OSH costs and OSH benefits are equalized at the margin. Furthermore, the set of tangencies between different firms’ offer curves and different workers’ indifference curves traces out an upward-sloping equilibrium wage-risk line. The remainder of this section is therefore devoted to reviewing the existing empirical evidence on, first, the determinants of the costs and benefits encountered by firms, and, second, the existence of CWDs in market economies.

3.1 Determinants of the OSH cost curve

A number of studies show that firms that take a more proactive stance toward the development of a comprehensive workplace risk prevention and assessment system are more likely to have lower accident rates than those which only follow the minimum legal requirements. The former firms may therefore benefit by having to pay lower wages as compensation for the exposure of employees to OSH risks.

Hunt and Habeck (1993) examined a sample of 220 firms in the US State of Michigan in order to establish the relationship between certain workplace risk prevention parameters and indices for the frequency and severity of accidents. The study suggests that there is a need for firms to generate and process internal information, investigate accidents and incidents fully, foster the emergence of a “prevention culture” and promote programmes to enhance workplace ergonomics. These findings are confirmed by Dembe et al. (2004), who suggest that targeted prevention strategies are required for reducing the likelihood of occupational injuries, such as worker self-assessment of the total physical effort demanded by a job and periodic monitoring of workforce job satisfaction.

Wilson (1996) shows further that health promotion programmes (e.g. ergonomic management, anti-smoking campaigns, purchase of personal protection equipment, stress management seminars, nutritional awareness) that take the form of comprehensive rather than single-goal programmes are most likely to succeed if there is appropriate coordination at all levels of an organisation (upper management, OSH professionals and personnel). Furthermore, in a study for Spain, Arocena et al. (2008) construct a risk prevention index that quantifies the intensity of firms’ preventive efforts, using a sample of 213 industrial firms. This measure is based on questions regarding six preventive dimensions, such as measures designed to eliminate risk at source, training, communication and workers’ participation, risk control, actions taken in view of foreseeable changes, documentation and emergency prevention, preparedness and response. Using a negative binomial regression that takes into account non-observable firm heterogeneity, it is shown that the intensity of occupational risk prevention is crucial for the reduction in the number of accidents. They also illustrate that there are important synergies between innovative preventive effort and organisational factors with respect to their effect on the improvement of OSH in a workplace.

The provision of OSH training to employees constitutes an additional strong preventative action that is associated with reduced workplace injuries and disease. In a wide-ranging literature review of published reports drawn from the period 1980 to 1996, Cohen and Colligan (1998, p. iv) find “overwhelming evidence to show the merits of training in increasing worker knowledge of job hazards, and in effecting safer work practices and other positive actions in a wide array of worksites”. The study shows that factors such as the size of the training group, the length/frequency of training, the method of instruction, the trainer credentials, and other extra-training factors (e.g. goal setting, feedback, motivational incentives and managerial actions) are significant determinants of the success of the training process. However, attention is drawn to the fact that many intervention studies fail to adequately decouple the provision of training from other forms of intervention (e.g. engineering, ergonomics).

Shannon et al. (1997) review the evidence on the relationship between the institutional structure of an organisation devoted to OSH promotion and injury rates. They focus on the influence of joint health and safety committees (e.g. representation, duties), the management style and culture (e.g. delegation of authority) and the organisational philosophy on OSH (e.g. role of top management, OSH training). They find that variables which are ‘consistently’ correlated with lower injury rates are those that reflect “a genuine concern by management for its workforce”, such as empowerment of the workforce, encouragement of long-term relations and systematic evaluation of safety hazards. No evidence is found...
to support the use of disciplinary procedures for safety violations or other types of policy regulation. Fenn and Ashby (2004) also argue that larger-sized firms may be more safety conscious given that they face a greater degree of monitoring by regulators, safety inspectors or insurers and/or that small firms may lack the necessary resources for investment in safe technology or health and safety programmes. The authors show that UK firms with a higher proportion of unionised employees and with health and safety committees are associated with a greater risk of reported injuries and illnesses (a finding that persists even after they correct for endogeneity). They suggest that this result is likely to reflect the superior reporting of accident and illness occurrences within establishments where the aforementioned internal governance mechanisms are present.

3.2 Determinants of the OSH benefit curve

Firms will endure the burden of engaging in costly OSH prevention and risk assessment efforts provided that the expected benefits outweigh the immediate outlays. Some of the most important benefits include lower rates of workplace injuries, illnesses, spells of sickness absence and a higher degree of overall job satisfaction on behalf of the workforce. Such features are desirable as they are likely to translate into a greater ability of firms to offer lower (‘compensating’) wages to employees and payment of lower insurance premiums or workers’ (sickness or post-injury) compensation. The literature summarised below identifies a number of important determinants which are related to the ability of firms to profit from the above benefits.

3.2.1 Accidents at Work

Gyekye and Salminen (2006) argue that the probability of work-related accidents has its roots in two major causes, namely the internal dispositional characteristics of workers and external causal factors such as the characteristics of the working environment. Dembe (2001) also takes into account the broader social, economic and cultural context.

In their attempt to identify the factors that are correlated with workplace injuries or accidents, researchers use a number of empirical methodologies. Poisson analysis (Alamgir et al., 2007), negative binomial models (Strong and Zimmerman, 2005; Blanch et al., 2009), ordered probit (Barling et al., 2003) and logistic regressions (Maiti and Bhattacherjee, 1999; Ghosh et al., 2004; Gauchard et al., 2006) are most commonly utilized. Askenazy (2006) also employs a bivariate probit model that takes into account the (non-random) process of reporting an injury by employees.

The literature suggests that there are significant variations in the rate of workplace injuries across individuals of different gender and distinct economic sectors. Men are most at risk from suffering an (predominantly fatal) accident (Krause et al., 2001; Askenazy, 2006). The manufacturing, construction, agriculture and transport sectors exhibit a higher injury rate compared to other industries. The incidence of occupational accidents is correlated with a number of other factors, such as company size (Fenn and Ashby, 2004), age (ageing and experience are negatively related to non-fatal accidents, though the reverse is true for fatal accidents) and outsourcing of labour (European Commission, 2009). Krause et al. (2001) and Dembe et al. (2004) show that there exists consistent evidence of a close link between occupational injuries and socioeconomic and work-related characteristics, such as low education, low family income, rural residence, unemployment history, blue-collar employment, long hours of work, monotony, lack of autonomy at work and job dissatisfaction. In particular, job satisfaction and other individual subjective evaluations of working conditions have also been used as predictors of the probability of work injury (Barling et al., 2003; Ghosh et al., 2004; Gyekye and Salminen, 2006; Gauchard et al., 2006). This research shows that job satisfaction may act as a mediating factor that lowers the chance of a work accident occurring by influencing workers’ commitment to safety management policies.

A number of studies show that there is a positive relationship between injury claim rates and precarious contract status. It has been suggested that particularly in times of higher aggregate demand there is no incentive for firms to train workers on temporary and casual contracts on matters of OSH (Guadalupe, 2003). Temporary workers are also unfamiliar with the workplace surroundings and with the operation of machinery and equipment, so they are more susceptible to workplace accidents. Probst and Brubaker (2001) show that employees with greater perceptions of job insecurity exhibit lower motivation and compliance to safety efforts. This, in turn, is related to higher levels of workplace injuries. Guadalupe (2003) shows that in Spain being in a fixed term contract increases the probability of an
accident by four to seven percentage points, which persists even after correcting for the systematic selection of workers into different contracts. However, Amuedo-Dorantes (2002) argues that the higher rate of work accidents observed in the case of temporary workers may be attributed to their inferior working conditions, because, once these are controlled for, fixed-term employees exhibit an even lower probability of suffering a work accident than their permanent counterparts. Garcia-Serrano et al. (2008) provide further evidence to support this hypothesis, using administrative register data from Temporary Help Agency (THAs) workers in Spain. They find that THAs have a lower probability of suffering from a serious/fatal accident and a smaller duration of absence compared to workers on non-THA temporary and open-ended contracts, after controlling for a set of personal, job and accident characteristics. They interpret this as an indication that agency workers potentially benefit from specific safety and health training programmes provided by THAs. Finally, Hernanz and Toharia (2006) analyse the effect of contract type on the rate of work-related accidents in Italy and Spain, using the 1999 ‘ad hoc module’ of the Labour Force Survey. They find that, once personal and job characteristics of workers are controlled for, there appears to be no difference in the probability of a work accident between open-ended or temporary contracts.

It has been argued that workplace accident rates exhibit a procyclical nature that is related to greater employee effort and stress, lack of experience, utilization of inadequate capital equipment or due to a greater prevalence of short term employment contracts during periods of economic boom (Kossoris, 1938; Brooker et al., 1995). Ashfaw et al. (2011) focus on industrial differences in the association between the business cycle and the incidence of workplace injuries in the US. Using nonfatal injury rates from 1976 through 2007 obtained by the Bureau of Labor Statistics database, their cyclical sensitivity is found to significantly differ across industries. Specifically, pro-cyclical associations between business cycle indicators and injury incidence are found in mining, construction and manufacturing but not agriculture or trade. The potential contributors to the positive cyclical association are also found to differ, with physical capital utilization being the primary culprit in the mining sector in contrast to labor utilization in construction. It is therefore argued that different prevention strategies are necessary in each sector. Crucially, additional precautionary measures appear to be required in these sensitive industries during periods of upturns in economic activity. Boone and van Ours (2006) investigate further whether the procyclical nature of workplace accident rates is truly related to a deterioration of workplace conditions. The authors suggest that high accident rates during periods of low unemployment arise as a spurious phenomenon caused by the reporting behaviour of workers. They propose that the reporting of an accident is likely to dent a worker’s reputation and raise the probability of lay-off. Hence one expects to observe lower reporting of accidents by workers in periods of high unemployment, when the likelihood and the costs of layoff are greater, and vice versa. Evidence in favour of such a mechanism is given by the fact that a negative relation is only found between unemployment and non-fatal accidents in 17 OECD countries. As no significant association is evident with respect to fatal injuries, the authors conclude that from a policy perspective there is no urgent need to worry about workplace safety when in cyclical upturns work-related accidents increase. A similar finding is also reported by Davies et al. (2009), who investigate cyclical fluctuations in the rates of workplace injury in the UK for the period 1986 to 2005. Their analysis demonstrates the importance of both compositional effects and individual reporting behaviour to understanding the variation of workplace injury rates across the business cycle. In particular, time-series analysis confirms that there is a positive elasticity of injury rates to fluctuations in economic activity that is stronger for minor injuries. Using a GLS fixed effects estimator to control for industry level effects, they also find that fluctuations in the rate of both minor and major injuries are related to the level of new hiring and the ratio of actual to usual hours worked. However, only minor injuries are found to be related to workers’ bargaining power. This suggests the presence of opportunistic behaviour on behalf of workers with respect to the reporting of minor injuries during times of favourable labour market conditions.

3.2.2 Occupational Diseases

In most developed countries it is recognized that illnesses are work-related provided that they are included in an official ‘list of occupational diseases’ by the national authorities. However, a general trend in recent years has been to move towards a more relaxed ‘open system’, where conditions of workplace risk exposure are considered on an individual basis (Walters, 2007, p. 17). Establishing the cause of work-related diseases is complex, as other non-work related factors may be responsible for causing an illness or aggravating a pre-existing medical condition. Furthermore, it may
take decades for some occupational diseases to develop and therefore to be listed in official registers (e.g. respiratory diseases such as asbestosis, silicosis and mesotheliomas).

The most prevalent health problems caused by work in the modern job market include musculoskeletal disorders, respiratory and skin diseases and psychosocial health problems (Krause et al., 2001). According to NIOSH (1997) and EU-OSHA (2000), musculoskeletal disorders (MSDs) are among the most common of work-related health problems in the US and the EU-27, predominantly affecting blue-collar workers. Around 25-30% of European employees are affected by them (e.g. 27% complain of backache and 23% of muscular pains), while they are estimated to account on average for 20.5% of the total compensation costs of European countries (Eurogip, 2004, p. 12). Respiratory and skin diseases are also quite prevalent occupational diseases, with exposure to asbestos, silica and other diseases of the respiratory tract accounting for more than 40% of total compensation costs in Europe, and skin diseases for an additional 10% (Eurogip, 2004, p. 12). Occupational asthma and dermatitis, in particular, are reported to have the greatest effects on productivity costs, ranging from 16.5 to 23 billion euros, respectively (Pickvance, 2005). Work organization factors such as job control, job demand, work pace and hierarchical relations are involved in provoking or aggravating psychosocial health problems (e.g. stress, depression, anxiety and pulmonary disorders) (Cox, 1978; Wege et al., 2008; Ali and Lindstrom, 2008). For example, EU-OSHA (2000) reports that around 30% of the European population is exposed to stress primarily caused by work, typically in service sector jobs. Joensuu and Lindstrom (2003) provide an extensive review of the role of stress and work factors in terms of affecting sickness absence. They report that in Sweden the percentage of long-term absentee due to psychological problems increased from 14% in the early 1990s to 25% in 2001. However, even though it is believed that psychosocial problems have an aetiological negative influence on health and well-being, it is difficult to provide an estimate of their overall economic cost. One potential reason for this is the interaction between physical features of the work environment with psychosocial elements. Leather et al. (2003), for instance, argue that there is a relationship between ambient noise levels and job stress, while Foppa and Noack (1996) show that stress at work is associated with musculoskeletal pain. More research is therefore required to detect the underlying causalities associated with psychosocial problems and attributes of the work environment.

Related to mental health outcomes is recent research that has examined the determinants of job satisfaction and its relation with individual health and overall well-being (Frey and Stutzer, 2002). Job satisfaction is consistently ranked as one of the most important facets of life satisfaction and of the quality of life (EPICURUS, 2007). It is also strongly correlated with aspects of employee behaviour and performance, such as absence or quits (Freeman, 1978; Clegg, 1983). Stansfeld et al. (1998) argue that job satisfaction may have an indirect influence on workers’ health through both physical and psychosocial avenues. Hence, improvements in job satisfaction over time may prevent workers from health deterioration. The meta-analysis of Faragher et al. (2005) provides a systematic and thorough review of the research linking work-related stress factors with ill health. Employee self-reported job satisfaction emerges as having by far the strongest link with overall wellbeing relative to other factors, while a strong positive relationship with both mental and physical health is found (both statistically and clinically). The relationships are particularly impressive for aspects of mental health, specifically burnout, lowered self-esteem, anxiety and depression, though the correlation with subjective physical illness is relatively modest. More recently, Fischer and Sousa-Poza (2009) employ objective measures of health from the German Socio-economic Panel (GSOEP) database. This is done in order to overcome the measurement error problems of self-reported assessments of health status and to address the issue of cross-sectional causality with job satisfaction. Importantly, the authors find that employees with higher job satisfaction levels feel healthier and are more satisfied with their health, so that improvements in job satisfaction over time have the potential to shield workers from adverse health developments. These results assume greater significance in the face of evidence suggesting that the intensification of work and poor job conditions in recent decades is associated with reduced employee motivation and job satisfaction (Green and Tsitsianis, 2005; Pouliakas and Theoossiou, 2009). Indeed, they highlight the possibility that a direct link might exist between OSH policies that improve conditions within the workplace and the state of employees’ overall physical and mental health.

3.2.3 Sickness Absence

The causes of absenteeism are in general multi-faceted, and are influenced not only by the health status of individuals, but also by the social insurance system, the work environment, biological factors, attitudes and commitment to work,
macroeconomic conditions and other social and psychological determinants. Regardless of the above complexities, the relevant literature distinguishes the causes of absence into two main components. On the one hand, it is viewed as the manifestation of workers’ labour supply decisions trading-off their limited input of time between the substitutable activities of ‘work’ and ‘leisure’ (Allen, 1981; Brown and Sessions, 1996). On the other hand, ill health and infirmity is believed to be a predominant factor underlying worker’s propensity to take days of leave (EUROFOUND, 1997, p. 21).

It is hard to obtain accurate measures of absence, especially since cross-country administrative data sources are relatively incomparable due to the marked differences in social insurance systems (in terms of the level of sickness benefits, length of time before payment, request of medical certificate inter alia). For this reason, a number of authors have attempted to measure the incidence and determinants of absenteeism via subjective measurements of lost work hours (Fenn and Ashby, 2004; Heywood et al., 2008; Pouliakas et al., 2010) or by exploiting the discrepancy in actual and usual hours of work from standard Labour Force Surveys (Barmby et al., 1991).

Applied psychologists were the first to study the causes of worker non-attendance, attributing its incidence primarily to the existence of job dissatisfaction (Steers and Rhodes, 1978). Nevertheless, some of the earliest economic studies, such as those by Barmby et al. (1991, 1995) in the UK, Ruser (1991) in the US, Johansson and Palme (1996) and Henrekson and Persson (2004) in Sweden, focus on the dynamic labour supply responses of employees to sick pay compensation or to other economic incentives that they face. These authors provide evidence of well-determined effects of the sick pay scheme and of the terms of the job contract (such as the wage) on the incidence and duration of absence. Other papers suggest that there are optimal reactions of employees to manipulations in their compensation via the provision of financial incentives (such as performance-related pay) (Wilson and Peel, 1991; Brown et al., 1999; Engellandt and Riphahn, 2004; Hassink and Koning, 2009; Pouliakas et al., 2010).

In addition, there appear to be cyclical fluctuations in the absence rate (Leigh, 1985; Kaivanto, 1997; Arai and Thoursie, 2005; Askildsen et al. 2005; Engstrom and Holmlund, 2007), since high unemployment is believed to act as a “discipline device” on the shirking behaviour of workers (Shapiro and Stiglitz, 1984). However, some researchers have emphasized that a potential “selection” mechanism is at work, whereby absence-prone workers are more likely to be fired in recessionary periods and (re-)hired during booms (Arai and Thoursie, 2005). Nevertheless, the fact that the strength of the procyclical nature of absenteeism persists even amongst countries with high employment protection legislation casts some doubt on the latter hypothesis (Lusinyan and Bonato, 2007).

Demographic factors are identified as important determinants of absenteeism. The female sickness absence rate is consistently higher than the male one (Barmby et al., 1991; Vistnes, 1997), a higher pattern of lost man-hours is observed with age (Ercolani, 2006), the presence of young kids in a household increase the likelihood of absence (Vistnes, 1997; Dione and Dostie, 2007) and the region of residence plays an important role (Barham and Begum, 2005). Furthermore, absence rates vary according to a plethora of job characteristics. Specifically, they decrease with rising earnings (Ercolani, 2006; Dione and Dostie, 2007) and increase with tenure, although absenteeism turns out to be low during probationary periods in which workers strive to impress their employers (Riphahn and Thalmaier, 1999; Ichino and Riphahn, 2005).

Industrial differences are notable, with industries with the highest absence rates typically found in the public sector (Ercolani, 2006). The proportion of employees on fixed term/temporary agency contracts, who face a greater risk of job loss, is negatively related to the absence rate (Hernanz and Toharia, 2006; Bradley et al., 2007), while a positive link between unionisation and absenteeism is also found (Leigh, 1981; Allen, 1984). A number of studies explicitly examine the correlation between the size of the firm and absenteeism (Barmby and Stephan, 2000; Fenn and Ashby, 2004). They show that absence rates are higher in larger-sized firms. This may imply that big firms face extra difficulties in monitoring their workforce, or that they are able to insure against absence via a buffer-stock of (substitute) workers.

Furthermore, the fact that the highest absence rates are observed in secondary sector occupations (e.g. process, plant and machine operatives), which are typically characterised by regular contact with hazardous materials and hard manual work, constitutes evidence of a correlation between absenteeism and poor working conditions. Ose (2005) shows that “bad” job attributes (high levels of noise in the work area, a high degree of monotonous work, heavy or frequent lifting or poor work postures) in the Norwegian labour market contribute to ill health, which, in turn, results in increased long-term absence. In a systematic literature review, Michie and Williams (2003) identify a number of job attributes as key factors associated with psychological ill health and sickness absence, such as long work hours, work overload, lack of control over work, lack of participation in decision-making, poor social support and unclear
management and work roles. Finally, Bokerman and Illmakunnas (2008) consider the interaction of a whole host of adverse working conditions variables, job dissatisfaction and sickness absence via a recursive multivariate probit model in the Finnish labour market context. They show that the prevalence of ‘harms’ at the workplace is associated with job dissatisfaction and dissatisfaction is, in turn, related to sickness absence. This implies that the improvement of working conditions should be considered a policy priority and an integral part of any policy scheme that is intended to decrease the incidence of absenteeism.

Variables reflecting the working time of individuals are often examined as covariates of the absence rate, since these have important implications for the mismatch between desired and contractual hours (Drago and Wooden, 1992; Brown and Sessions, 1996). Other factors that are examined include whether employees engage in shift work, whether they are on part-time contracts or if they are entitled to any other form of flexible working time arrangements (e.g. annualised hours, flexitime, working from home). Lusinyan and Bonato (2007) argue that flexible working time arrangements are generally found to be associated with lower employee absence because they reflect the extent to which the constraint of contractual hours is relaxed. However, it is important to bear in mind that firms which find absence to be less costly might also be those that offer flexible working time schedules in the first place, so estimates of the effect of non-binding hours on absence usually suffer from endogeneity bias. Moreover, shift work and irregular working time patterns are found to interfere with the biological and social rhythm of employees (Finn, 1981; Dahlgren et al, 2006), so they are believed to contribute to higher absence.7

Recent research has focused explicitly on the internal workings of organizations, in an attempt to identify the equilibrium rate of absenteeism within a given labour market. Coles et al. (2007) use a unique matched employer-employee dataset from France, showing that the nature of the production technology of firms can play a significant role in terms of the shadow price for absenteeism (i.e. the amount that firms are willing to pay to achieve a unit reduction in the absence rate). Specifically, they argue that absence is more costly in firms employing a so-called “assembly” or “just-in-time” production technology. The latter are therefore more likely to respond to absence via greater monitoring and the provision of additional and/or more generous sick pay. Coles et al. (2007) augment their analysis with the supply-side of the market, by identifying the increase in pay that is required by workers for a marginal reduction in their hours of absence. In this manner, they obtain structural estimates of the equilibrium rate of absence.

In a similar spirit, Pauly et al. (2002), Heywood and Jirjahn (2004) and Heywood et al. (2008) focus on the interrelation between teamwork, monitoring and absence. It is confirmed using representative samples of German and UK establishments that firms with interdependent productivity (team production) face a higher cost of absence, and as a consequence expend additional resources on the task of monitoring absence. Firms with teamwork subsequently face a lower absence rate relative to those that do not rely on productive interactions among workers.

The difficulty of obtaining internationally comparable data on worker absenteeism has hampered any attempts to analyze cross-country differences in patterns. Barnby et al. (2002) highlight that similar forces operate across nine advanced Western economies, such as a negative relation of absence with the wage rate. Bergendorff et al. (2004), Osterkamp and Rohn (2007) and Lusinyan and Bonato (2007) investigate the relationship between international sickness absence rates and institutional characteristics of labour markets, after controlling for differences in age structures, health, unemployment and participation rates (among other determinants). They find higher absence rates in the Netherlands, Norway, Sweden and the UK relative to other countries. They also confirm that the level of sick pay and the generosity in terms of granting sick leave are important determinants of cross-country differences in sickness absenteeism. Moreover, absenteeism is found to be significantly and positively related to the strictness of employment protection (EPL) that increases the cost of dismissal to employers. Nonetheless, this finding is disputed by Frick and Malo (2008), who claim that once the effect of EPL on absenteeism is disentangled from the effect of the type of contract, there is no significant association between EPL and the number of absence days.

Finally, it is worth mentioning that a number of researchers have recently turned their attention to the determinants and consequences of the related phenomenon of sickness presenteeism, defined as the tendency of workers to attend employment despite illness (Chatterji and Tilley, 2002; Skatun, 2003; Dew et al., 2005). Though research on this issue is still in its infancy, concerns have been raised given evidence that sickness presenteeism is related to future sickness absenteeism (Aronsson et al., 2000; Bergstrom et al., 2009; Bockerman and Laukkanen, 2010), and that the cost of the former in terms of lost productivity may be substantially greater than that of the latter (Goetz et al., 2004).
3.3 Empirical Evidence on Compensating Wage Differentials in Market Equilibrium

3.3.1 The Hedonic Wage Methodology

There exists an extensive literature providing estimates of the presence and extent of CWDs via estimation of so-called *hedonic wage equations*. Predominant attention has been paid to the “*value of a statistical life*” approach (Thaler and Rosen, 1976), which attempts to detect the monetary value that employees themselves place on their lives, on the basis that they are willing to accept jobs of a higher likelihood of a *fatal accident* occurring.

Acknowledging that *non-fatal accidents and illnesses* can also have debilitating and life-changing effects, economists have widened the scope of the value of life approach by implementing a hedonic wage framework that also considers non-fatal risks. Conventionally, a (linear) Mincer-type wage equation is estimated in an attempt to describe the market equilibrium line $EE$ in Figure 2, as follows:  

$$
\ln w_i = \beta_0 + \beta_1 f_i + \beta_2 r_i + x_i \gamma + u_i
$$

(1)

where $w_i$ are earnings of individual $i$ $(i = 1, ..., N)$, $f_i$ and $r_i$ are fatal and non-fatal job-related risks, respectively, the vector $x$ contains a number of confounding factors reflecting individual and job characteristics that are used as control variables and $u_i$ is a random error term. In various specifications the vector $x$ has also been specified to include the workers’ compensation benefits that are payable for job injury (Butler, 1983; Viscusi and Moore, 1987). The regression estimates of the parameters $\beta_1$ and $\beta_2$ are hence believed to provide an unbiased measure of the wage premium required to induce workers to accept an extra marginal unit of (fatal or non-fatal) job risk. The parameter $\beta_1$ has also been typically used to calculate the *value of a statistical life* (VSL), namely the total monetary amount that a society would be willing to pay in order to reduce the probability of death by one individual.

Numerous studies suggest that, *ceteris paribus*, workers in jobs associated with a higher degree of fatal risk and unhealthy working conditions command a higher wage compared to safer jobs (Thaler and Rosen, 1976; Olson, 1981; Marin and Psacharopoulos, 1982; McNabb, 1989; Viscusi, 1993; Arabsheibani and Marin, 2000; Viscusi and Aldy, 2003). In particular, the majority of the estimates of VSL are significant and exhibit considerable variation around the $3$-$9$ million range (Viscusi, 1993, p. 1930; Viscusi and Aldy, 2003). This relatively wide discrepancy arises mainly due to differences in the samples examined and alternative empirical specifications. For instance, Viscusi (1993, p.1928) notes that some of the earlier studies tended to use data pertaining to high risk occupations (Thaler and Rosen, 1976). This resulted in underestimation of VSL due to self-selection of individuals with lower wage-risk preferences into hazardous jobs. Researchers have also employed appropriate interaction terms of the fatal risk variable with a number of other factors to explore potential heterogeneity in CWDs. For example, Thaler and Rosen (1976) and Aldy and Viscusi (2007) use interaction effects with age, since one should expect age-related differences in the propensity of workers to accept job risk (possibly related to variation in consumption patterns over the lifecycle). Another source of bias in CWDs is the endogenous choice of union membership, since wage levels are restricted by collective bargaining. While a number of studies have found evidence in favour of the hypothesis of the endogeneity of union membership (Marin and Psacharopoulos, 1982; Siebert and Wei, 1994), others have failed to establish a convincing pattern (Arabsheibani and Marin, 2000). The evidence of gender disparities in wage-risk tradeoffs is also ambiguous, which is likely to be due to the greater concentration of women in white-collar jobs with low fatality risks (Ruser and Butler, 2009, p. 308).

Although the majority of studies confirm that there exist positive and significant CWDs for *non-fatal injuries* (Viscusi and Aldy, 2003), a number of authors have highlighted that such wage-risk tradeoffs do not always transpire (Siebert and Wei, 1994; Sandy and Elliott, 2005; Wei, 2007). These studies have typically suffered from an inability to distinguish between alternative types of severity, frequency and duration of injuries. In general, estimated CWDs are found to be larger for relatively more severe injuries that cause lost workdays (Viscusi, 1993, p. 1935).

Sandy and Elliott (2005) and Wei (2007) investigate the existence of CWDs for job-related *illnesses*, highlighting the fact that employees face a far greater risk of illness at work rather than fatal or non-fatal *injury*. Sandy and Elliott (2005) find evidence of significant CWDs for manual male workers only. Wei (2007) confirms this finding using data from the UK’s 1998 Workplace Employee Relations Survey (WERS), which permits the construction of establishment-
specific illness and accident rates. He finds that the estimated wage compensation for one job-related illness episode per year ranges from 27% to 140% of annual earnings, depending on gender and estimation methodology.

Notwithstanding the appeal of the aforementioned approach, a number of researchers argue that there are significant shortcomings associated with the estimation of hedonic wage functions. The most prominent of these are measurement error in the risk variables and the endogeneity between job risk and earnings (Moore and Viscusi, 1990). The former is said to arise due to “aggregation bias” (Lalive, 2003), given that broad industry or occupation level measures of job risk are typically used in the literature. However, this masks the important discrepancies in the level of job-specific risk faced by people who perform different tasks, even if they are classified within the same industry or occupational category. The risk data also fails to distinguish between injuries that are attributable to the job environment and those which are due to workers’ safety behaviour (Viscusi, 1993, p.1925). Another source of measurement error may also be at work in studies that ignore the influence of non-fatal job risks on workers’ remuneration, since the latter are collinear with respective fatal risks (Arabsheibani and Marin, 2000). In addition, many occupational diseases are underrepresented in official government registers, while existing data sources typically contain insufficient information on specific non-fatal risks (Karnon et al., 2005). Elliott and Sandy (1998) illustrate that overstatement of job disamenities by employees who are dissatisfied with their pay brings into question the validity of CWDs obtained on the basis of subjective employee responses to questions about job risk. Following the above arguments, Viscusi (2004) confirms that measurement error bias tends to lead to underestimation of estimates of CWDs.

The heterogeneity in individual preferences for risk and income is likely to compromise estimates of CWDs too, by masking important unobserved factors that can simultaneously affect both individuals’ taste for earnings and their preferences for job risk (Garen, 1988; Hwang et al., 1992; Hintermann et al., 2008). For instance, it is often argued that since safety can be considered to be a normal good, people in non-manual jobs or those who enjoy a higher level of human capital can “afford” to select jobs with better working conditions (Biddle and Zarkin, 1988). On the demand side it might also be the case that employers offer superior market opportunities (both pecuniary and non-pecuniary) to high-skilled workers, due to job-specific training of the latter (Viscusi, 1993, p. 1917). The above arguments imply that the wage-risk tradeoff is likely to be an increasing function of wealth, with the “best jobs” in society also being the “best paid”. This highlights the need for empirical studies to carefully control for the heterogeneity in firm and employee characteristics related to ability, education and productivity. Proposed methods to circumvent this problem in the literature have included the use of either rich matched employer-employee datasets, panel data methods that control for unobserved individual (but not firm) heterogeneity or utilization of appropriate instrumental variable (IV) techniques (Arabsheibani and Marin, 2000; Johansson and Palme, 2001; Wei, 2007; Hintermann et al., 2008). Nevertheless, a significant shortcoming of the longitudinal and matched employer-employee studies is their inability to accurately capture heterogeneity in technologies and product/factor prices, which are underlying determinants of the market price for risk. The IV approach also suffers from an ad hoc choice of instruments.

3.3.2 The Stated Preferences Approach: Willingness to pay for job safety

In light of the above deficiencies of the revealed preferences approach, recent research has employed the stated preferences methodology of estimating the “willingness-to-pay” (WTP) of individuals for risk reduction. Subjective WTP estimates typically utilize survey methods in order to elicit the magnitude of the rate at which individuals are willing to forgo income in exchange for additional safety, which would nevertheless keep their utility unchanged i.e.

\[ \text{WTP} = \frac{\partial W_i}{\partial r_i} \]  

(2)

where in a perfectly competitive labour market and pertinent to a local range of risk increments, it is expected that \( WTP = W'(R) = \beta_2 \), where \( \beta_2 \) is the estimated coefficient of the hedonic wage equation (1).

The stated preferences methodology has become increasingly popular among economists since the early 1990’s, despite the strong reservations of mainstream theorists who traditionally rely on “what people do” rather than on what they “say they will do” (Easterlin, 2004). Indeed, the longstanding tradition in economics is to infer decision processes...
on the basis of observed choices in market settings, and to treat any subjective data as suspect due to the incomparability and inconsistency of individual responses (Samuelson, 1938). However, powerful theoretical and practical reasons have been advocated in favour of eliciting subjective evaluations of expectations and events by survey respondents. It is argued that due to the unobservability of counterfactual outcomes, researchers often act with partial information. Therefore, observed choices may be consistent with many alternative specifications of preferences and expectations. It is therefore likely that revealed preferences do not necessarily conform to the rational expectations hypothesis of axiomatic expected utility theory (Manski, 2004, p. 1330). Indeed, behavioural research (described in detail in Section 4.1. below) emphasizes that informational biases, cognitive inadequacies, impulsive behaviour and time inconsistency are underlying human traits that may drive a wedge between intended welfare and actual market choices (Thaler and Sunstein, 2008). On this basis, the revealed utility concept is argued to be inadequate for the purposes of public policy evaluation as it essentially assumes the optimality of individual behaviour (Kahneman et al., 1997). In addition, WTP estimates allow the assessment of the relative efficacy of various health and safety interventions that entail non-marginal risk changes, as opposed to hedonic estimates which focus on risk decisions at the margin. On a practical level, the need to employ stated choice data is also spurred by the unavailability of necessary information for evaluating non-market goods, such as natural resources (Hanley, 1989; Adamowicz et al. 1994; Hanley et al., 2001) and health care (Ryan and Farrar, 2000). Empirical estimation of stated choice models is encouraged further by the significant number of studies over the last thirty years indicating that “practically speaking, stated and revealed preferences seem to match up surprisingly well in different choice contexts, cultures and time periods” (Louviere et al., 2000, p. 12).

Stated preferences research was originally carried out by using the so-called contingent valuation approach. This approach relies on asking respondents to provide direct monetary valuations of alternative policy measures on simulated market contexts (Gerking et al., 1988; Miller et al., 2002). To overcome the inconsistency of simple open-ended questions, which are subject to cognitive burden and response bias (e.g. protest bids, yeah saying, embedding effects), researchers have developed more pragmatic contingent markets (Diamond and Hausman, 1994). This is achieved via the design of a realistic payment vehicle (such as the specification of sensible income-taxation trade-offs or comparisons of goods in a similar nonmonetary dimension) and by providing sufficient information for individuals to make informed choices.

Originally used in the fields of marketing research and transport (Luce and Tukey, 1964; Green and Srinivasan, 1978), the complementary choice modelling approaches of conjoint analysis and of relative valuations have been increasingly employed by economists to enable valuation of multidimensional policies. Both of these methods seek to imitate the actual choices of individuals in quasi-market settings by asking them to make a series of deliberate hypothetical choices between alternative scenarios (which are differentiated by attributes and levels). In the conjoint analysis approach, the WTP of individuals is indirectly obtained by including an income variable within the set of attributes describing a realistic job scenario (or ‘vignette’). Van Beek et al. (1997), for instance, use such a method to detect employers’ preferences for desirable characteristics of job applicants, van Leeuwen and van Praag (2002) assess policy measures for stimulating lifelong on-the-job training in the Netherlands and Pouliakas and Theodossiou (2010[b]) report estimates associated with workers’ WTP to avoid insecure job contracts and physically demanding jobs. In the relative valuations approach, instead, indirect monetary valuations are obtained against a suitable peg event for which a reliable valuation exists. Karnon et al. (2005), for instance, evaluate the benefits of the prevention of non-fatal work-related accidents and ill health pegged against monetary valuations for the prevention of road deaths.

Stated preference methodologies constitute a valuable contribution to informing policymakers about the effectiveness of novel and multidimensional OSH interventions, since they allow researchers to widen the scope of risk options available to respondents that may not be available in market data. Caution is nonetheless warranted on the grounds that choices elicited in hypothetical settings are likely to be affected by the degree of ‘contextual realism’ that the researcher establishes for the respondents (Dolan and Kahneman, 2008, p. 225). In particular, the set of attributes need to be free of errors in risk perceptions, in the sense that they should closely correspond to probability events that individuals encounter in reality and which can be carefully evaluated on the basis of pre-existing experiences (Viscusi, 1993, p. 1942). The processing of risk information and the undertaking of complex choices by individuals within a survey context is also reliant on their cognitive capacity (learning and fatigue effects) and on their commitment to the exercise (Manski, 2004, p. 1343). Finally, stated choices must obey similar axioms such as those stipulated in rational
consumer choice theory (e.g. transitivity and monotonicity), as individuals may resort to the use of heuristics (e.g. maximin strategies) in the face of complex scenarios (Hanley et al., 2001, p. 449).

4. Sub-optimal Allocation of Job Market Risk

The existence of CWDs that are conducive to the attainment of an efficient degree of OSH provision on behalf of firms hinges critically on the strong requirement that competitive markets forces are at work. Nevertheless, convincing arguments exist on both efficiency and equity grounds that predict that the equilibrium level of OSH provision, as determined in the theoretical realm of the free market in Figure 2, is likely to be suboptimal.

4.1 Efficiency considerations

The efficiency of market forces, as shown by point \( S_f^* \) Figure 1, is likely to be diluted due to ineffective collective bargaining and transaction costs, important externalities in the provision of OSH, moral hazard problems in compensation insurance markets, asymmetric information and systematic biases in individual risk perceptions. As Figure 3 makes clear, the above market inefficiencies imply that the competitive level of OSH provision might turn out to be less than the efficient level (\( S_f^* < S_f^* \)). In this case, a discrepancy arises between society’s willingness to pay for risk reduction and market valuations of the cost of OSH when measured at current levels of risk exposure (Biddle and Zarkin, 1988; Herzog and Schlottmann, 1990).\(^9\)

4.1.1 Externalities

If the contractual parties do not internalize all of the damage costs of workplace injuries or diseases in their OSH decisions, the social costs of risk reduction will exceed the marginal costs faced by individual agents. In this case, economic theory predicts that a less than optimal degree of OSH is likely to prevail in the economy.

For instance, within the workplace it may be the case that reckless behaviour on behalf of some workers may lead to a greater probability of injuries for their colleagues (Viscusi, 2006, p. 19). Similarly, despite the offer of workers’ compensation insurance (discussed in more detail below), it is unclear whether firms fully account for the effects of unfortunate OSH occurrences on the continuing employment displacement, disablement and quality of life of employees (Weil, 2001). Job injuries/illnesses are likely to hamper the work ability and ex post productivity of workers following an incident. The working time lost during a recovery period may also have negative implications for the workers’ stock of human capital and their subsequent earning capabilities (Woock, 2009). This might explain why individuals who experience injuries and diseases related to work face a higher probability of unemployment, engage in early exit from the labour market or encounter increased difficulties with respect to re-entering into suitable employment (Pransky et al., 2005). Crichton et al. (2005) examine the effects of work-related accidents on the future labour market outcomes of individuals, using data from New Zealand. They highlight that injury severity (defined as those cases where individuals receive earnings compensation for more than 3 months) is associated with negative future prospects. They show that injured individuals of longer duration have lower employment rates and incomes after returning to the labour force, with women, older-aged individuals and the lower-paid being more susceptible to negative outcomes. Similar findings are provided by Reville and Schoeni (2001), who report that lower employment opportunities and increased income losses are observed for workers with permanent partial disability due to occupational factors. Workers in the manufacturing sector seem to suffer from the greatest economic losses when compared to the rest of the workforce, although this loss is ameliorated when the severity of injury is smaller. It is therefore proposed that national policies should have a dual purpose, namely to provide economic support to such target groups via adequate compensation benefits and also to improve prevention programs that will increase OSH. Woock (2009) finds that the annual earnings losses are highest for workers who suffer from a work-limiting disability following injury. In line with the above, a recent study estimates the impact of bad health on individual productivity and employment status in the UK (HSE, 2008). Using the standard Mincer (1974) specification of an earnings equation on data from the British Household Panel Survey (BHPS), it is
found that an individual’s health has a strong and significant impact on productivity. Specifically, those who are in excellent health earn 4-7% more than those with average health, while poor health suffer from a 7-15% wage penalty. Moreover, the probability of being in work is found to decline as the individual’s health status worsens, with those benefiting from excellent (poor) health having a 17% (34%) higher (lower) probability of being in employment compared to those with average health.

Retirement decisions are particularly affected by OSH episodes. Newton et al. (2007) provide evidence that a significant portion of the European labour force remains idle following an adverse OSH event, as individuals do not feel capable of performing the work that they did prior to the incident. Pransky et al. (2005) argue that 11% of the workers in their sample planned to retire earlier due to the experience of a work-related injury, with job dissatisfaction acting as a significant mediator in the occupational illness-retirement nexus. Tüchsen et al. (2009) confirm that job injuries are a strong determinant of the decision of individuals to take a disability pension retirement, although this is only found to hold for their male sample.

In addition to intensifying the strain on public finances and pension systems, there is a dearth of empirical research on the impact of poor health and workplace accidents on macroeconomic and social outcomes. In a study commissioned by the HSE (2008) in the UK, the effect on GDP of ill health caused by work is explored based on the estimation of a regional growth equation that covers the period 1995 and 2005. The analysis shows that a higher incidence of poor health and long-standing illness is associated with negative economic growth, with an elasticity of 0.02. This effect is relatively robust to the inclusion of controls for industry structure. Moreover, Dembe (2001) highlights that in addition to the more traditional costs of occupational injuries, the social consequences of accidents and ill health on the lives and daily activities of affected workers, their families and their immediate communities are sizeable and have generally been neglected in the public discourse.

The debate about the external consequences of accidents and illnesses is reflected in discussions of the appropriate legal compensation for future wage losses in a tort suit. Indeed, it is often argued that “tort compensation for personal injuries generally provides for economic damages and coverage for noneconomic loss that is below the levels of the value of statistical life pertinent for establishing efficient incentives for accident prevention” (Viscusi, 2006, p. 26). Lewis et al. (2003) and Butt et al. (2008) have shown that the current method of awarding damages in tort in UK courts, which relies on actuarily-determined labour market hazards (known as Ogden Tables), fails to compensate injured claimants fully for their loss of future earnings. This is found to be particularly the case for a non-uniform group of claimants, namely men, younger individuals, the disabled and those with post-injury earnings capacity. The source of the under-compensation is attributed to the fact that the current multiplier-multiplicand method used to settle injury claims, which converts annual losses in earnings due to injury into a discounted lifetime capital sum, neglects to take into account earnings growth over a claimant’s working life and the ‘conditionality’ of future employment patterns on current activity status.

4.1.2 Moral hazard in insurance markets

Workers’ compensation insurance (including sickness and income benefits) and tort law provide strong incentives that can influence the behaviour of workers and firms with respect to observed injury and claim incidence, duration of absence and costs (Ruser and Butler, 2009, p. 315). On the employee side, safety precautions taken on the job, filing claims for injuries sustained and the duration of time taken off work following an occupational accident or illness may all be influenced by the anticipation of ex post compensation. Similarly, liability for compensation of injured workers and the associated payment of insurance premiums may act as deterrents that induce firms to respond by either investing in OSH improvements or by engaging in manipulative reporting and/or lobbying efforts.

In most advanced industrial countries employers are generally responsible by law to contribute towards insurance schemes that cover for the medical expenses of injured workers, as well as the replacement of a portion of the lost stream of future earnings in case of employment displacement and continuing disablement. In some cases firms are also required to provide for rehabilitation services and the payment of income benefits to dependants in the case of workplace fatality (Ruser and Butler, 2009, p. 313). A survey of the costs and funding of injury and disease compensation schemes in Continental Europe finds significant variations among countries, where the cost of occupational diseases accounts for over three quarters of the total cost of compensation and exceeds that of occupational injuries by around 10-20%.

Table
1 shows that the cost to insurance organizations for compensation and rehabilitation in 2000 ranged from over 1 billion euros in Germany and Italy to 23 million in Austria (Eurogip, 2004, p.6). In addition, it has been estimated that the cost of occupational injury, measured in terms of the compensation paid by insurance companies to victims who brought civil liability suits against their employers, is between 100 to 250 million euros per year (Eurogip, 2004, p. 5). However, the extent to which a prospective reduction in these damage costs has a sufficient bearing on an employer’s safety performance depends on the criteria for pricing insurance and on employer-employee ‘risk-bearing’ and ‘claims-reporting’ moral hazard (Butler and Worrall, 1991; Shapiro, 1999, p.603).

4.1.3 Asymmetric information and systematic biases in risk perception and preferences

Any positive impact that higher indemnity benefits might have by inducing firms to undertake safety improvements, are likely to be mitigated by the fact that workers may compromise their vigilance for safety and increase the frequency of (fraudulent?) claims. As summarized by Ruser and Butler (2009, p. 325), the available empirical evidence generally concludes that rising benefits are associated with an increase in the incidence of claims (with a claim rate/benefit elasticity that ranges between 0.4 and 1.1), particularly for harder-to-diagnose cases (e.g. back sprains). There is also evidence of a positive link between the magnitude of benefits and the duration of claims. Meyer et al. (1995), for example, compare claims duration using a “natural experiment” setting of differential maximum income benefits offered by two US states. They find that time out-of-work spells increased for those eligible for higher benefits, though the value of their benefits/duration elasticity of 0.4 is smaller compared to other estimates in the literature (Ruser and Butler, 2009, p. 331).

In many countries the insurance premiums paid by firms are not a real reflection of their true safety record. In some cases (e.g. US), firms’ contributions differ with respect to the degree to which the premiums are “experience-rated”, with smaller-sized firms often paying standardised “manual rates” calculated with respect to an average level or risk within given industrial-occupational classifications. In other states (e.g. Sweden, Austria, Portugal, Belgium) a single rate is applied to all enterprises regardless of activity and risks and usually as a percentage of the payroll (Walters, 2007, p. 21). The divergence that is caused between the price of workers’ compensation insurance and the actual OSH conduct of most employers is therefore expected to weaken their incentives to invest in safety, since this will have a negligible effect on their future insurance premiums. Empirical evidence on whether such an “experience-rating effect” prevails broadly supports this hypothesis, though not all research has been conclusive (Ruser and Butler, 2009, p. 326).

Employers may also be inclined to engage in other methods to avoid reparation payments, which are cheaper than OSH prevention efforts. Some examples include the exertion of pressure on workers by either discouraging them to file a claim or to misreport its severity, placing them on light duty or expediting a premature return to work. Companies also often aggressively pursue lobbying activities that seek to influence policymakers with respect to the stipulation of rules governing worker’s compensation insurance (e.g. capping legal fees, making types of injuries noncompensable, restricting the choice of doctors etc.). Such actions often result in employers not bearing the full cost of workplace accidents. It is also notable that firms are often held accountable for only a small portion of occupational illnesses, given the difficulties in establishing causation between the workplace and occupational diseases and in proving workers’ exposure to the latter (Shapiro, 1999, p. 608). All of the above avenues may serve to alleviate the attainment of an efficient degree of OSH prevention on behalf of firms.
workers face in a job. This will result in lower wage premiums demanded per unit of actual risk, as they are not entirely aware of the dangers to which they will be exposed to in their employment.

Evidence from cognitive psychology and behavioural economics research also highlights that even if employees are fully informed about objective probabilities pertaining to job hazards, systematic biases in individual risk perception are likely to result in consistent miscalculation of the odds of accidents/illnesses at work occurring. According to the predictions of prospect theory (Kahneman and Tversky, 1979), in the presence of uncertainty people do not always process messages about risk in a ‘rational’ manner (Kahneman et al., 1982; Slovic, 2000; Kahneman and Tversky, 2000). There is considerable evidence that human beings have an inbuilt resistance to unwelcome information and that they suffer from cognitive limitations (Denscombe, 1993). For these reasons they tend to employ simplifying heuristics in judgement and decision making i.e. simple rules of thumb which enable individuals to handle the enormous amount of information at hand by engaging in simpler operations (Tversky and Kahneman, 1974; Gilovich et al., 2002). However, these heuristics sometimes lead to systematic errors in peoples’ interpretation of risk and in their subsequent choices, which may depart from those expected on the basis of perfect rationality.

A number of these biases have been described in the literature. Tversky and Kahneman (1974) identify the so-called representativeness heuristic, whereby people tend to under-use base-rate information in forming judgments. This leads to misperceived probability assessments that either over-infer patterns from short sequences of events that depart from the norm, or exaggerate random lengthy streaks (a phenomenon that Rabin (1998, p.24) calls the “law of small numbers”). Another important heuristic includes anchoring and adjustment, which focuses on the fact that arbitrary starting points may serve as an anchor for further probability assessments, while adjustments in assessments away from this anchor are typically insufficient. Furthermore, the availability heuristic posits that individuals are influenced in their assessment of the likelihood of risks by pre-existing, recent or readily available experiences (Thaler and Sunstein, 2008, p. 27).

The availability heuristic may account for the fact that people often overestimate the frequency of rare yet dramatic or sensational causes of death (e.g. homicide) and underestimate the frequency of less well-publicised yet common causes (e.g. stroke and asthma) (Slovic et al., 1982). For instance, in an experiment that the prominent psychologist Amos Tversky carried out on 120 Stanford graduates, who were asked to assess the probability of dying from various causes, the estimate given by the group regarding the likelihood of an accident occurring was 0.32, which differed significantly from the ‘true’ statistical estimate of 0.05 (Bernstein, 1996). This divergence between subjective and objective perceptions of risk by individuals is also often exacerbated by the fact that human beings are not very receptive to or tend to misread any new information that contradicts their initially formed hypotheses, known as confirmatory bias (Rabin, 1998, p. 26). Another pervasive attribute includes optimism bias, since people are commonly found to overestimate their personal immunity from harm, such as the chance of experiencing an occupational accident or illness (Weinstein, 1989). Interestingly, even though it is often argued that learning and adaptive behaviour on behalf of individuals may act to moderate or ameliorate the above errors in individual judgment (Viscusi, 2000), evidence from psychological experiments have failed to support the above notion (Rabin, 1998, p. 31).

In addition to any possible misperceptions in the evaluation of workplace risk, there is also contention about the economic assumption that workers have stable and well-defined preferences. Rabin (2000) has proved that conventional expected utility theory fails to provide a plausible account of risk aversion over moderate stakes, as a simple calibrating exercise shows that assumed risk aversion over small bets implies a rapidly diminishing marginal utility for wealth. Instead, it is proposed that a prominent reason for why people behave in a risk-averse manner with regard to small risks (even though expected utility theory suggests that they should behave in an approximately risk-neutral manner) is loss aversion, the tendency of individuals to value the pain of losses more acutely than the pleasure of same-sized gains (Tversky and Kahneman, 1991; Rabin and Thaler, 2001, p. 226). Loss aversion has also been considered to be a culprit for the fact that the marginal willingness to pay for risk reduction from an accustomed danger level is typically observed to be less than the financial inducements required by individuals to accept incremental additions to risk of a similar magnitude. A related explanation focuses on the so-called status quo bias of prospect theory, which asserts that people are likely to prefer their given state of affairs and overreact to the presence of newly identified risks (Rabin, 1998, p. 14). In general, workers’ wage-risk tradeoffs are likely to depend on whether the risk levels deviate substantially from individuals’ “reference levels” of risk, which may be determined by previous job experiences or family background circumstances inter alia. For instance, the required CWDs of employees might be greater for risk levels that are close to
their reference points rather than for those that are substantially further away (a phenomenon known as diminished sensitivity bias) (Rabin, 1998, p. 15). All of the above behavioural arguments have important implications about the curvature and continuity of the constant utility loci shown in Figure 2, and are therefore expected to result in wage-risk tradeoffs that are different compared to the standard predictions of the economic paradigm.

Overall there is scepticism in the literature about whether the choices individuals make in response to objective information sets regarding wage-risk bundles in various jobs are such that maximize their well-being in the first place. Gilbert (2006, p.109) has summarized the evidence from numerous psychological experiments, which point to the fact that people make systematic errors in their perception of the future. One prominent reason is that they fail to account for the fact that current emotions intrude on assessments of the future ("presentism"). Individuals are also quite poor in forecasting how quickly hedonic adaptation takes place, whereby changing reference points in response to unfortunate events of life and trauma (e.g. paraplegia, unemployment) result in consistent overestimation of the negative impact of the latter on future well-being (Brickman et al., 1978; Frederick and Lowenstein, 1999; Dolan and Kahneman, 2008; Pouliakas and Theodossiou, 2010[b]). Biases have also been detected in people’s own assessment of their future utility on the basis of recollections of past memories and choices, with peak and end effects obscuring their own evaluations. In addition, there is tenuous evidence regarding the stability and dynamic consistency of people’s preferences. Framing effects are known to lead to different elicited preferences when decision makers choose among two presumably equivalent options that are nonetheless stated in a different manner (Tversky and Kahneman, 1986). People are also sensitive to context effects and preference reversals when faced with a menu of choices (Rabbin, 1998, p. 37-38). Finally, due to individuals’ propensity towards immediate gratification, their current choices are often inconsistent with their long-term preferences (Thaler and Susnstein, 2008, p. 44). All of the above behavioural arguments would therefore suggest that there is a discrepancy between individuals’ ex-ante “decision utility” and their ex-post degree of welfare or “experienced utility” (Kahneman et al, 1997).

4.2 Equity considerations

It is often pointed out that important equity considerations arise in the market for OSH, due to the fact that the social class distribution of occupational accidents and illnesses is skewed. Indeed, job risk tends to affect a relatively vulnerable portion of the workforce (e.g. lower educated/lower income individuals) to a far greater extent than individuals in higher rungs of the social strata (Henderson, 1983). Crucially, this may justify public regulation of OSH levels even among employees who receive the same CWD within a given occupational category, since there may still be inequity in the spread of occupational accidents and diseases across individuals who belong to the same broad employment group.

CWDs are also only likely to arise in a free market when alternative employment opportunities are open to workers, as the mobility of labour will, in theory, spur competition among employers in terms of providing an attractive [wage, safety] bundle. However, in industries where accident or illness risk is most prevalent (e.g. manufacturing, construction, transportation) there is a higher incidence of workers suffering from a low pay-no pay vicious cycle (Asplund et al., 1998; Stewart and Swaffield, 1999; European Commission, 2004; Stewart, 2007). Jobs are normally offered on a take-it-or-leave-it basis and refusal of the job offer from the part of the worker may imply unemployment or another dead-end job as the only outside options available. Under such circumstances of labour market segregation, employees are expected to be reluctant to quit their current job. Discrimination may also account for the fact that different labour market groups (e.g. of different race or ethnicity) may face quite different offer curves and thus settle in different hedonic market equilibria, with the group facing larger risk being compensated less (other things equal).

Furthermore, the recent growth in precarious employment in advanced market economies has been generally accompanied by the erosion of coverage of workers’ compensation and legal entitlements in case of injury or disease, a weakening of their bargaining power in terms of filling claims and less emphasis on rehabilitation strategies on behalf of employers (Walters, 2007, p. 8). Important equity considerations therefore arise with respect to the inequality of opportunity faced by workers who tend to cluster in the lower-end of the labour market (Pouliakas and Theodossiou, 2010[a]).
The literature suggests that labour unions can play a potentially significant role with respect to safeguarding equity in the market for OSH (Beaumont, 1983; Walters, 1995). Labour unions use their political influence, collective bargaining and participation in health and safety committees to influence workplace health and safety standards (Hirsch et al., 1997; Morse et al., 2003; Donado, 2007). However, the empirical evidence is not conclusive regarding the effects of unions on OSH. Freeman (2004) suggests that trade unions contribute to the improvement of working conditions and of compensation benefits for employees who suffer from work-related health problems. He argues that, in general, unions effectively represent employees’ interests with regards to health and safety at the workplace. Litwin (2000) finds that the presence of unions in the British labour market contributes to a reduction of job injuries. Similarly, Siebert and Wei (1994) show that unionised workers experience lower fatal injury rates. Reilly et al. (1995) also illustrate that lower injury rates are observed in plants with union representatives in health and safety committees. In contrast, Fishback (1986), Fenn and Ashby (2004) and Donado (2007) provide evidence suggesting that increased unionisation is associated with a higher incidence of workplace injuries and illnesses. This conflicting evidence may be an outcome of endogeneity, since it is not only the case that unionisation affects OSH but the reverse might also be true. In particular, hazardous working conditions may motivate workers to participate in unions and seek for voice strategies in order to protect themselves (Robinson, 1990; Hirsch and Berger, 2001; Fenn and Ashby, 2004; Nichols et al., 2007).

5. Government Intervention in the OSH Market

Given the market inadequacies described in the previous section, government intervention might be necessary for an economy to attain efficient and equitable levels of OSH. Such government activities typically take the form of the provision of a social security safety net to employees (especially in countries that do not rely on a private insurance market), elimination of informational deficiencies, the setting of standards and the imposition of financial penalties or prosecution to non-compliers. The extent to which such actions are beneficial, though, in the sense that they improve on market outcomes and enhance societal welfare, needs to be assessed rather than assumed (Viscusi, 2006).11

5.1 Social security

In most developed countries legally mandated systems exist for providing benefits to workers who suffer from occupational disease or injury. Though in several cases private insurance associations are developed that operate on their own or in parallel with a public social security scheme, it is typical for the state alone to administer compensation for harm in many countries. Adema and Ladaaque (2009) report that in the year 2005 approximately 2.6% of the GDP of the OECD group of countries was devoted on average to expenditure on Incapacity Benefits (including care services, disability benefits and pensions, benefits accruing for occupational injuries and diseases and employee sickness payments). Significant variations by country are observed mainly due to differences in the criteria for recognition of occupational diseases and eligibility for compensation, with Mexico (0.1%), Turkey (0.2%) and Korea (0.6%) lying at the lower end of the spectrum, the US (1.3%), the UK (2.4%) and the EU-19 group of countries (3%) found in the middle, and the Scandinavian countries [Sweden (5.6%), Norway (4.4%) and Denmark (4.3%)] featuring at the top (OECD, 2009).

Significant cross-country variation also exists in the institutional features governing the provision of occupational disease and injury compensation schemes. In several countries employers are responsible for full coverage of their employees’ salary for an initial period (typically 3 days), after which benefits from the social insurance scheme take over. Although the basis for compensation traditionally is to recompense victims for loss of earnings capacity, some countries (e.g. Denmark, Finland, Sweden) take into account the damage caused on long-term physical and mental function and diminished quality of life (Walters, 2007). In most countries a medical assessment of illness or injury is required, which can range from mere validation of disability to the pursuit of a more “aggressive” requirement of job incapacity evaluation (e.g. the Pathways to Work initiatives of the UK government). Moreover, there are significant discrepancies in qualifying periods, replacement rates (ranging from 50 percent in Austria to the total of the ceiling earnings in Luxembourg and Finland), and in the maximum duration of benefit provision (Walters, 2007, p.43-46).
5.2 Information disclosure and indicators of OSH

When the source of market failure lies in the discrepancy between the perceived and actual risks faced by employees, information disclosure by the government could ameliorate the problem, provided that it is clear, well-organized and with sufficiently new information content that avoids information overload (Viscusi, 2006, p. 27-28). Improving workers’ access to information about the nature, severity and relative employment risks that they face is likely to improve market performance, by empowering them to form accurate perceptions of the required wage-risk trade-off. Providing appropriate training to workers with respect to evaluating the typically overwhelming amount of risk information that is available to them may also serve to tackle any systematic biases in hazard assessment as described in section 4.1.3 (Shapiro, 1999, p. 601).

Recently, policymakers (especially within the US and EU) have sought to construct appropriate measures or indicators of health and safety that will inform the decisions of major stakeholders within their countries. The purpose of such indicators is to quantify the state of affairs in the workplace infrastructure, inputs and outputs of individual economies. As Rantanen et al. (2001, p. 17) suggests, the socio-economic structure of the economy, including demographic conditions, the industrial or service-oriented nature of markets and the technological frontier, are crucial factors for determining the state of workplace inputs and the exposure of different risk groups of employees to them. Exposure to work-related risk factors is, in turn, one of the main determinants of health and safety outcomes, such as the absence behaviour of employees (partly mediated by their job satisfaction), the incidence of work accidents and the occurrence of occupational diseases. These negative OSH outcomes are important as they entail significant direct and indirect economic costs at both an individual and societal level.

A number of important research projects have been undertaken under the auspices of major international organizations such as the WHO, ILO and the EU, with the purpose of designing appropriate OSH indicators. These projects identify a plethora of generic and specific indices of OSH, covering a wide range of OSH contingencies. Primary examples of these projects include the European Community Health Indicators Project (ECHI) (Kramers, 2003), the assessment of the state of OSH throughout the EU Member States by the EU-OSHA (2000), or the European Commission’s emphasis on the creation of ‘good work’ (European Commission, 2001; EUROFOUND, 2007). The above echoes the ILO’s preoccupation with the attainment of ‘Decent Work’ (ILO, 2005). Other important contributions include those of the Finish Institute of Occupational Health (FIOH) (Rantanen et al., 2001), the US National Occupational Research Agenda (NORA) project (CSTE, 2005; CDC, 2007) and the EU’s WORKHEALTH and HEALTHatWORK projects (Kreis and Bodeker, 2004; HEALTHatWORK, 2010). The latter projects develop a comprehensive set of OSH indicators that become operational by separating them into clear-cut policy domains, such as “OSH policy and infrastructure”, “OSH hazards/work conditions” and “OSH outcomes”. Table 2 below summarises some of the most commonly cited indicators of OSH, which have been arranged in a manner that corresponds to the structure of this survey.

5.3 OSH regulation, enforcement and compliance

There are important differences in the OSH acts and regulations of countries, which usually focus on the criteria required for the receipt and duration of sickness/disability benefits, rules determining the framework under which workers’ compensation and tort liability operate and the conduct of OSH inspection and enforcement. The ILO has a central role in providing recommendations and guidance for national OSH policies. Although the ILO publishes instruments, which outline varying levels of obligations for member countries, the latter have to ratify each instrument before implementation takes place. Ratification does not guarantee implementation. The recommendations are for national governments to implement or use as a guiding policy.

A number of studies investigate the factors that underlie the decision to ratify and adopt OSH-related ILO conventions. The length of ILO membership, national income status and regional affiliation are shown to be associated with a higher number of ratifications by member states (Wilson et al., 2006). Boockman et al. (2009) focus on developing countries, for which OSH is seldom given priority status due to its costly nature. They employ an empirical methodology that allows for the fact that ratification behaviour is influenced by unobserved characteristics both of
countries and of different conventions. The presence of such effects is argued to stem from the fact that some conventions may be more easily ratified than others, since they differ in terms of their degree of flexibility or complexity. Finally, other factors that are identified as possible obstacles to the ratification process include the lack of infrastructure, the lack of political will, incompatibility with national legal systems and peer effects (Chau and Kanbur, 2001).

The adoption of government regulations in many countries is often governed by principles that are unrelated to society’s willingness to pay for risk reduction, such as the need to take a precautionary stance against dimly understood risks using an individual risk threshold or a “technology-based” approach (Shapiro, 1999, p. 612; Viscusi, 2006, p. 39). This is in contrast to the usual economic methodology of assessing the efficiency of regulations by utilizing a cost-benefit approach, where the discounted benefit stream, quantified in terms of the value of a statistical life, must exceed the cost per life saved. The latter is argued to constitute the most cost-effective method of manipulating the population mean risk level and thus making the best use of society’s scarce resources (Sunstein, 1990), although important objections are expressed on the grounds that cost-benefit analysis is unreliable and unethical (Shapiro, 1999, p. 618). Furthermore, it is pointed out that significant caution needs to be exercised when undertaking stringent regulatory actions, given that offsetting behavioural consequences (moral hazard) may negate any beneficial impacts of the regulation. For instance, the positive effect of any legislation forcing employers to supply personal protective equipment (PPE) to their employees might be mitigated by the subsequent complacency of workers, thus leading to more workplace accidents. This is particularly likely to be the case when people overestimate the efficacy of a policy. The overall assessment of the efficacy of regulations is also complicated by the fact that there might be substantial opportunity costs associated with the diversion of resources away from other potentially health-enhancing expenditures (e.g. better health care) (Viscusi, 2006, p. 61).

The success of OSH regulation is also dependent on enforcement by the relevant authorities of the stipulated OSH standards and compliance on behalf of the affected agents. Because compliance with a regulation sometimes involves substantial administrative and other costs, rational firms will choose to comply provided that the expected costs of noncompliance are greater. Given that the latter are a function of the likelihood of detection and the magnitude of the expected penalty levied, it follows that a low probability of detection plus low fines for violations of OSH standards are likely to compromise the efficiency of regulation (Ruser and Butler, 2009, p. 334). Empirical evidence tends to suggest that the estimated effects of OSH inspections on safety are quite small or non-existent (Shapiro, 1999, p. 613), although research in this area has found it difficult to circumvent the problem of the endogenous nature of injuries and inspections (since inspections usually take place at high-risk industries).

Several recent studies examine the effectiveness of OSH regulations in promoting OSH labour standards. Wilson et al. (2007) assess the relationship between the ratification status of OSH-related ILO conventions and the reported occupational fatality rates of ILO member countries, after controlling for possible confounding factors (such as income levels and length of ILO membership). They show that non-ratifying countries generally have statistically higher fatal injury rates. Based on this evidence, they argue that all countries should promote ratification of ILO conventions aimed at improving OSH conditions. Arocena and Nunez (2009) also highlight the importance of OSH legislation in reducing workplace accidents in the “advanced” manufacturing sector (i.e. industries utilising a high degree of technology and skilled workforce) in Spain. However, they fail to identify any significant change in the incidence of accidents in traditional manufacturing sectors following the adoption of new OSH legislation.

A notable study of the UK’s Health and Safety Executive (HSE, 2006) examines the relationship between the OSH activity of firms and sectoral performance, using a gross output growth accounting framework. This research is important as it addresses the question about whether the adoption of costly OSH regulations are likely to compromise the performance of firms and have negative effects on their economic activity. It utilises Cambridge Econometrics’ annual time series of input-output consistent volume and value indices of change for 42 sectors of the UK economy from 1970 and the Bank of England Industry Dataset (2003). Moreover, it incorporates data on OSH activity and stringency measures in the UK collected by the HSE. No clear evidence is provided on the impact of OSH stringency on productivity, although it is suggested that the former might have a negative short-term effect on output as firms struggle to meet stricter regulatory requirements. However, in the long-run, the so-called Porter hypothesis (Porter, 1991) asserts that newer and better technologies are likely to be adopted as a response to regulatory incentives, thereby enhancing productivity over time. Given that the latter positive effects may only become apparent after a substantial time lag, it is
argued by the authors that future research needs to tackle the insufficiency of time-series data which inhibits the identification of a time trend on the impact of regulation stringency data on productivity.

6. Policy Recommendations and Issues of Future Concern

The focus of public policy on OSH has shifted considerably in advanced market economies during the past two decades, due to the need to respond to the new challenges posed by rising service sector jobs, adverse demographic developments and constraints in public finances. From a strict focus on the prevention of occupational injuries and industrial diseases, OSH has now become a more encompassing concept that considers the overall protection of workers’ health via prevention, maintenance of working capacity, compensation, rehabilitation and, in general, the fostering of an OSH culture. This survey highlights a number of issues that have direct application to public policy. In order to address some of the inefficiencies in the market for OSH, policymakers can engage in various actions that can either influence the incentives that workers and employers face or employ direct regulatory activities. Governments can ensure that workers have adequate access to information about job hazards and can require from employers to provide adequate training to their employees in order to mitigate potential biases in the processing of such information. They can also influence the cost of noncompliance to firms by manipulating the criteria (e.g. eligibility, coverage etc.) set in workers’ social insurance schemes for offering compensation to injured/ill workers. Mandating that insurance premiums reflect more accurately the health and safety records of enterprises is another avenue of action. However, there has been distress by employers over what are claimed to be exorbitant premiums paid, combined with perceptions of unfairness by workers regarding settlements that presumably underestimate the true cost of harm. This has resulted in a slight reorientation in recent years away from ‘no fault’ compensation schemes and towards civil law litigation in several countries (Walters, 2007, p 37). Furthermore, the authorities can enforce safety standards via targeted inspection, large penalty fines and regular monitoring of OSH indicators. However, calls for employers to bear a Pigouvian injury or illness tax have failed to materialize due to the difficulties in setting appropriate tax rates that reflect the incidence of occupational diseases and due to foreseen tax avoidance problems (Shapiro, 1999, p. 610).

Notwithstanding the greater availability of data sources on health and safety at work and the large number of studies, a number of relevant topics on OSH require further exploration in future research. It is expected on theoretical grounds that the need for OSH regulation is likely to decline over the long-run reflecting improvements in safety technology, greater substitutability of capital for labour and higher incomes that raise the demand for safety (Ruser and Butler, 2009, p. 347). However, it is unclear whether the underlying causes of the falling accident rates observed in advanced market economies in recent years can be attributed to the aforementioned mechanisms, or whether they are related to exportation of dangerous activities to Third World countries or an increase in informal employment (EU-OSHA, 2005).

The implications of the demands of modern job markets for OSH are not adequately researched. The rising incidence of precarious employment raises concerns about the inadequacies in social insurance coverage, weakened claims due to inefficient bargaining power and insufficient training and reintegration efforts for workers on non-indefinite contracts. The fact that many of the traditional “industrial” OSH risks (e.g. injuries, respiratory illnesses) are now aggravated by the increasing stress levels, MSDs and mental disorders associated with job insecurity and market flexibility draws attention to the fact that the causes and effects of multifactorial risks need to be investigated further. Similarly, the changing demographic evolution of the workforce towards an increasing proportion of female, racially diverse and older-aged employees implies that future studies need to investigate closely the OSH risks posed to these groups. There is gender discrimination in social insurance, for instance, caused by the non-inclusion of many female-dominated occupations in the official lists of registered occupational diseases (Walters, 2007, p. 31). Migrants and ethnically or racially diverse workers are a particular cause of concern, since such individuals are typically employed in high-risk and informal sectors, they face linguistic and cultural barriers that prevent appropriate OSH communication and training and they are often not covered by social security systems (HSE, 2004). Indeed, given that a substantial part of the working population is employed under the vulnerable conditions of the informal economy, often possessing low levels of skills and experiencing low/irregular incomes and long working hours, extension of OSH measures to include such informal workers is a challenge for the future (ILO, 2008, p. 19).
The effectiveness of OSH rehabilitation and reintegration strategies of firms and national insurance providers, which are likely to become progressively important in the face of the increasing strain on social insurance systems, needs to be studied further. It is not yet clear what types of rehabilitation and reintegration procedures (e.g. training partially disabled workers in new job tasks, adapting workplaces etc.) are more effective than others. Furthermore, a related issue of concern is the greater vulnerability of Small and Medium-Sized Enterprises (SMEs) in terms of providing adequate OSH prevention and rehabilitation policies given the dearer credit-constraints that they face.

Due to absence of appropriate longitudinal data sets very few studies examine the labour market prospects of employees that return to work following long-term periods of sickness absence or disability, or the overall (economic and non-economic) consequences of such prolonged periods of inactivity on individual performance and welfare. In general, studies that attempt to identify the causal relationship between work factors, individuals’ health status and economic and welfare outcomes are fraught with problems of endogeneity, subjectivity bias and measurement error. There is also scarcity of empirical research on the causes and consequences of occupational diseases, though improved datasets are likely to allow for more accurate measurement of their incidence and identification of correlations with individual and job characteristics in the future.

The harmonisation and international comparability of appropriately-designed OSH indicators in terms of comparable definitions, data collection methodologies and quality also remains one of the major challenges of forthcoming endeavours. Most governments in the developed world are presently seeking to redefine the traditional remit of OSH towards taking a more forward-looking and comprehensive stance that allows anticipation of the major threats posed by modern labour market practices. Nevertheless, such a task is hindered by the substantial differences in social laws and administrative practices across countries which cannot be easily incorporated in the usual statistical methodologies.

Finally, in both policymaking and academic circles there has been a reluctance to embark on economic evaluations (cost-benefit analysis) of regulatory policies related to OSH (Viscusi, 2006). Any methodological and ethical obstacles that inhibit accurate economic evaluations of such policies are nevertheless likely to be overcome in the future, given the overall need to address issues of affordability and efficiency in social security funds and workers’ compensation schemes. Overall the jury is still out in terms of whether the apparent ineffectiveness of OSH enforcement is simply a reflection of adequate market incentives contributing to the realization of an almost-optimal equilibrium level of OSH. It is expected, though, that the fostering of an inclusive culture of OSH via cooperation of all of the major stakeholders of an economy (workers, enterprises and governments), has the potential to complement the forces of the market and imply significant long-term efficiencies.

Acknowledgments
This research forms part of the HEALTHatWORK project (An inquiry into health and safety at work: A European Union perspective) - a project supported by the European Commission through the 7th Framework Programme "THEME [HEALTH-2007-4.2-3] Grant agreement no: 200716. We are grateful for the suggestions and comments of two anonymous referees of this journal.

References
Aronsson, G., Gustafsson, K. and Dallner, M., (2000), ”Sick but yet at work: An empirical stud


Figure 1 Efficiency in the Market for OSH

Figure 2 Equilibrium in the Market for OSH
Figure 3 Inefficiency in the Market of OSH

Table 1 Cost to insurance organizations for compensation and rehabilitation in seven European countries 2000
### Table 2  Selected Measures and Indicators of OSH

<table>
<thead>
<tr>
<th>OSH Measures</th>
<th>Examples of OSH monitoring indicators</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. OSH Costs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OSH Budget</strong></td>
<td>% of firms’ budget devoted to OSH promotion, prevention and rehabilitation</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td><strong>OSH Training</strong></td>
<td>% of workforce received OSH training within firms</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td></td>
<td>% of firms’ budget devoted to OSH training activities</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td><strong>OSH Management</strong></td>
<td>% of enterprises that engage in OSH risk assessments</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td></td>
<td>% of enterprises that keep records of work-related accidents/illnesses/sickness absence</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td></td>
<td>% of enterprises that have official OSH policies and infrastructures in place (e.g. joint worker-manager committees to discuss OSH-related issues and frequency of meetings; OSH employee representative; targets for OSH performance)</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td></td>
<td>% of enterprises that have arrangements in place to support the return to work of employees on long-term sickness absence/illness/disability</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td><strong>2. OSH Benefits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Presence of Physical Agents in the Workplace</strong></td>
<td>% employed exposed to high levels of noise, vibration, radiation, abnormal room temperature etc.</td>
<td>FIOH</td>
</tr>
<tr>
<td><strong>Presence of Chemical Agents in the Workplace</strong></td>
<td>% employed exposed to/in regular contact with dangerous products or substances (e.g. asbestos, lead, benzene, pesticides etc.)</td>
<td>FIOH</td>
</tr>
<tr>
<td><strong>Ergonomic Conditions in the Workplace</strong></td>
<td>% employed whose job involves inconvenient work postures, repetitive movements, lifting of heavy loads etc</td>
<td>FIOH</td>
</tr>
<tr>
<td><strong>Working time arrangements</strong></td>
<td>% employed working very long hours (at least 50h/week) or in irregular shifts (especially nights shifts)</td>
<td>FIOH</td>
</tr>
<tr>
<td><strong>Psychosocial Problems</strong></td>
<td>% employed who report mental health problems (stress, anxiety, depression, burnout)</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td></td>
<td>% employed subjected to threat of physical violence/harassment/threats/violence at work</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td><strong>Outputs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fatal work accidents</strong></td>
<td>Incidence of fatal accidents at work per 100,000 employees</td>
<td>FIOH WORKHEALTH</td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
<td>Source(s)</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td><strong>Non-fatal work accidents</strong></td>
<td>Incidence of non-fatal accidents at work with more than 3 days of absence per 100,000 employees</td>
<td>FIOH WORKHEALTH</td>
</tr>
<tr>
<td><strong>Occupational illnesses</strong></td>
<td>No. of ill health cases related to a recognized occupational disease by economic activity per 100,000 employees</td>
<td>FIOH WORKHEALTH</td>
</tr>
<tr>
<td></td>
<td>No. of ill health cases related to a recognized occupational disease per 100,000 employees who were working in the last 12 months</td>
<td>HSE</td>
</tr>
<tr>
<td><strong>Sickness absence</strong></td>
<td>% of usual working hours lost in a reference week due to absence related to sickness or ill health</td>
<td>Barmby <em>et al.</em> (1991)</td>
</tr>
<tr>
<td></td>
<td>% of employed absent from work in reference week due to own illness, injury or temporary disability</td>
<td>WORKHEALTH</td>
</tr>
<tr>
<td><strong>Job Satisfaction / Job Quality</strong></td>
<td>Mean trend of subjective employee well-being scores by country and facets of jobs (e.g. pay, quality of working conditions, job security, working hours etc.)</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td><strong>3. OSH Regulation</strong></td>
<td>% of GDP devoted to expenditure on incapacity benefits, including sickness allowances and disability benefit</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td></td>
<td>Coverage of workers’ compensation system (% of total number employed)</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td><strong>The Size of the Social Insurance System</strong></td>
<td>Ratification rate of relevant ILO conventions on OSH</td>
<td>FIOH</td>
</tr>
<tr>
<td></td>
<td>Proportion of human resources devoted to labour safety inspection and enforcement (either as part of national health and safety executives or in workplaces) (inspectors per 1000 employees)</td>
<td>FIOH</td>
</tr>
<tr>
<td></td>
<td>% of OSH inspections that are prosecuted/convicted</td>
<td>HEALTHatWORK</td>
</tr>
<tr>
<td><strong>Enforcement and Compliance with OSH Activities</strong></td>
<td>Annual workers’ compensation benefits paid and mean amount paid per covered worker</td>
<td>NIOSH</td>
</tr>
<tr>
<td></td>
<td>Differences in participation probabilities between formerly injured/ill/disabled employees and healthy employees</td>
<td>HSE HEALTHatWORK</td>
</tr>
<tr>
<td></td>
<td>Differences in wage outcomes between formerly injured/ill/disabled employees and healthy employees</td>
<td>HSE HEALTHatWORK</td>
</tr>
</tbody>
</table>

*Notes: FIOH = Finnish Institute of Occupational Health (*Rantatten et al.*, 2001); HSE = UK Health and Safety Executive; WORKHEALTH (*Kreis and Bodeker*, 2004); NIOSH = National Institute of Occupational Safety and Health (*CDC*, 2007)*
Endnotes

1 For example, disability and sick-pay account for nearly a quarter of social-welfare expenditure in Norway, which spends more than 24 times as much on these payments as it does on unemployment insurance. In Britain, unemployment benefits account for less than a tenth of what disability and sickness expenditures cost to the government budget.

2 For example, the rate of non-fatal and fatal accidents has exhibited a significant downward trend (of 27.4% and 42.4%, respectively) in the last decade in the countries of the EU (European Commission, 2009).

3 It needs to be noted that there is a burgeoning literature examining aspects of the health situation of the population in general, such as the effect of various pathogenic behaviours (e.g. smoking, diet, lifestyle) on mortality. However, this survey focuses on issues related to the factor ‘work’ and examines the implications of the latter for individual (and societal) health and safety.

4 “The wages of labour vary with the ease or hardship, the cleanliness or dirtiness, the honourableness or dishonourableness of the employment” (Smith, 1776, Book 1, Chapter X, Part 1).

5 Deleire and Levy (2004) in the US and Grazier and Sloane (2008) in the UK show that individuals exhibiting strong aversion to risk (which is proxied by family structure) make occupational choices that are biased towards the selection of safer jobs.

6 This has been typically attributed to the differential out-of-work commitments of the two genders (e.g. domestic duties such as the rearing of children). Ichino and Moretti (2006), instead, show that a significant fraction of the male-female absenteeism gap can be explained by a 28-day cycle, which vanishes for workers aged 45 or older. They interpret this as evidence that the menstrual cycle is responsible for the higher level of female absenteeism.

7 However, Tuchsen et al. (2008) fail to find any conclusive evidence of a link between shift work and absenteeism in a sample of Danish workplaces, after controlling for an extensive list of relevant factors (such as age, education, body mass index, smoking status, alcohol consumption, leisure time physical activity and other psychosocial and physical work environment factors).

8 Viscusi (1993, p. 1916) also describes the use of a more complex structural equation systems approach, which has relied on the estimation of two-equations, a utility locus and a market opportunity curve. This approach relies on making specific assumptions about the functional form of the utility function, and has used regional economic variables as appropriate identifying constraints of the demand side equation.

9 Evidence that workers’ willingness to pay for risk reduction (i.e. the slope of their indifference curves) exceeds the implicit market-clearing price (i.e. the slope of the isoprofit curves of firms) has been obtained with the use of “job switching” regression estimates. The latter exploit the fact that workers “vote with their feet” among jobs in various industries, which differ with respect to working conditions and wage compensations. Combining these with estimates from conventional hedonic wage equations, one can predict the wage-risk tradeoff at the “quit margin”.

10 While in some continental European countries it is possible to claim industrial injury or disease benefits by recourse to civil litigation, in most cases social security bears the burden of compensation. For the purposes of ensuring social peace, some countries (e.g. US, Germany) have also adopted the so-called ‘no-fault’ systems approach, whereby employers on suspicion of causing harm are not liable for civil action as long as they adhere to their legally specified contributions to the workers’ compensation scheme (Walters, 2007, p. 15).

11 Viscusi (2006) offers a comprehensive discussion regarding the fact that many government agencies (e.g. EPA) tend to use upper bound estimates of the component parameters used in calculating risk. They also focus on exposure levels with “reasonable certainty of no harm” (the precautionary principle) and emphasize dimly understood risks. This conservatism bias tends to yield risk assessments that are at the upper tail of the actual risk distribution, which is in contrast to the normative approach of assessing economic benefits based on society’s total willingness to pay for risk reduction.