Estimating individual occupational risk using registration data

P.H.G. Berkhout & M. Damen
RIGO Research Consultants, Amsterdam, The Netherlands

ABSTRACT: In this paper the occupational risk of individual employed workers is assessed using merged registration data. The analysis features a negative binomial count data regression model for statistical inference. In the model the hazard rate is determined by variables characterizing the personal background and working conditions, such as age, sex, tenure, contract type, working hours and sector. The estimation results can be used to assess individual occupational risk and identify high-risk groups and branches of industry. The main contribution of the paper to the literature is the fact that calculations include the entire population at risk and are based solely on registration data on individuals, jobs, firms and accident casualties.

1 INTRODUCTION

Risk assessment is a cornerstone of the European approach to prevent occupational accidents and ill health. European legislation with respect to risk assessment is grounded on the Framework Directive 89/391, which has been transposed into national legislation of member states. The European Agency for Safety and Health at Work (EU-OSHA) suggests a stepwise approach to risk assessment: 1) identifying hazards and those at risk; 2) evaluating and prioritizing risks; 3) deciding on preventive action; 4) taking action and 5) monitoring and reviewing. Clearly, the process of risk assessment builds on identifying hazards and the workers at risk in the initial step. This paper sets out to quantify the occupational risk of all registered employed workers, based on merged registration data and conventional statistical methods.

Usually statistics on occupational accidents are presented on the aggregate level. Risk assessment of individual workers is complicated, since it requires detailed information on outcomes of all workers at risk, not only the accident casualties. Information gathering is usually initiated when occupational accidents occur; the counterfactual - workers who under the same circumstances did not get involved in an accident - is not observed. For instance, Rivas et al (2011) analyze with the help of data-mining techniques 62 questionnaires to be completed whenever an accident occurred. They conclude that data-mining outperform conventional statistics in terms of the predictive function and the possibility of identifying interactions between variables with a bearing on accidents. However, the question is whether inferences drawn from a sample of accidents only can be extrapolated to the workers who did not get involved in an accident. In experimental lingo this is referred to as the external validity, see Shadish, Cook & Campbell (2002). In sample surveys of the population at risk generalization is usually not a problem. But sample surveys are usually too small in relation to the probability of observing enough accident casualties in the survey sample.

In the present paper we will demonstrate that plausible estimations of individual occupational risk can be extracted from detailed registration data representing the entire population at risk. The negative binomial count data model is used to estimate the individual rate of accident occurrence per working day. To this end complete registration data on casualties of serious occupational accidents was merged to individual registration data describing the entire population at risk. In the model occupational risk may vary for individual workers as a result of differences in their human capital (age, gender, flex worker, nativity, tenure, hours worked per week) and the hazardous nature of the work (instrumented by 262 sector fixed-effects covering the entire economy).

Since the available data cover a time span of thirteen years (1999-2011), the analysis is not only focused on cross-sectional explanation of occupational risk. In addition, time trends and business-cycle effects were identified. A declining number of accident casualties in this decade was in fact one of main reasons why this study was initiated. Estimation results show a decline of 27% in the average rate of accident occurrence in time. This raises the question whether occupational safety in The Netherlands has truly improved to such extent. We conjecture that structural changes in labor market supply and demand have contributed significantly to changes in the observed average accident rate. For one, ageing of the worker population pushes the average rate upward (since occupational risk tends to rise with age for workers older than 30). The size of this ageing effect is considerable, due to an increase in the average age of the worker population from 36.3 in 1999 to 39.5 in 2011. On the other hand, a hazard-biased destruction of jobs over time has changed the composition of the job-pool in favor of the less hazardous jobs. A lower average rate of accident occurrence would be the result. Roughly estimated this composition-effect accounts for one third of the total
decline. All in all, a ‘true’ decline in the accident rate of 19% in 13 years remains, implying a considerable improvement in occupational safety in The Netherlands in the past decade.

The paper is structured as follows. In Section 2 the observed individual occupational risk is considered within the conceptual framework of labor market matching of workers to jobs. Data, definitions and a formal description of the empirical model are presented in sections 3 and 4. In Section 5 estimation results are presented. Section 6 concludes.

2 CONCEPTUAL FRAMEWORK

2.1 Labor market matching of individual skills to job requirements

In this paper the rate of occurrence of occupational accidents – the accident frequency – is analyzed within the human capital framework of labor market matching of workers to jobs. On the labor market the skills of individual workers are matched to jobs requiring certain skills. Each worker-job match results in a hazard rate specific to that particular match.

Each existing job represents an objective hazard potential intrinsic to the job. This potential is not affected by the worker selected for the job. The job hazard potential is an unobserved (latent) variable in the framework. One might say that in case an accident occurs, part of the potential’s hazardous energy is released and observed. The hazards require skills to handle them. These hazard handling skills are supplied by the workers. Anything contributing to the ability of handling hazards successfully can be considered: vitality, experience, responsiveness, knowledge etc. We refer to the aggregate all of such characteristics as the ‘hazard handling skills’. In the framework of matching workers to jobs these skills can be treated as any other requirement for the job.

Workers are heterogeneous in their skills endowment, jobs are heterogeneous in their hazard potential. Since occupational accidents involve costs (and accident prevention reduces these costs), the employer treats the worker’s hazard handling skills as any productive skill. Depending on relative prices, there’s a trade-off between the cost evading hazard handling skills on the one hand and productive skills incurring revenue on the other hand. Should accidents incur high expected costs employers will emphasize hazard handling skills in the selection of workers. Vice versa, hazard handling skills will have low value to the employer if expected costs of accidents are low. Expected costs are defined by actual costs multiplied by the accident probability. Given a worker-job match, the job’s latent hazard potential is transformed into an observed occupational risk. Thus from the worker’s perspective his individual occupational risk will vary with the job he is matched to. From the employer’s perspective the occupational risk related to his jobs varies with the matched workers.

In the present analysis the individual occupational risk is defined as the rate of accident occurrence per working day. For the sake of brevity we will refer to the rate of accident occurrence as the ‘hazard rate’. This concept is commonly used in survival analysis, defining the rate at which individuals leave a certain state, conditional on having survived up to that point. In the present context we might say workers ‘survive’ as long as they are not involved in an occupational accident. Note that the term ‘hazard’ does not refer to the actual occupational hazards (i.e. risks or dangers) the worker is confronted with in his job. Instead, it refers to the risk of ‘not surviving’, that is of getting involved in an accident. The hazard rate is estimated on the individual level. Hazard handling skills are instrumented by age, sex, nativity, hours worked, tenure and type of contract. On the demand side fixed-effects for 262 sectors of industry are estimated, representing job average hazard potentials.

2.2 Structural changes in labor supply and demand

The above matching framework is used to explain changes in the population average over time. That is, developments on the macro level are related to mechanisms operating on the micro level of matching workers to jobs. We will assume matching technology did not change during the thirteen years of observation. With the matching technology given and fixed relative prices of hazard handling skills and productive skills, two different ways to change the average accident frequency on the demand side of the market are identified. First, the safety conditions in existing jobs may be affected by technological progress (e.g. improved gear and materials), investments in the firm’s safety culture and regulation by authorities. These factors affect the hazard potential in existing jobs. One might say they are associated with the ‘intrinsic’ or real occupational safety. Second, on the demand side there is on-going dynamic process of destruction and creation of jobs. This process in itself may be hazard-biased in a sense that it is selective with respect to hazard potential. When for instance due to technological progress relatively hazardous job are destroyed and relatively safer job are created, the composition of the job-pool changes over time. All else held constant, this will be revealed in a declining average hazard rate. However, such a composition-effect can also contain a spurious improvement of occupational safety. For example, when relatively hazardous jobs are shifted to self-employed individuals the average hazard potential in the entire job population declines, but occupational safety from the national perspective remains unaltered (the hazardous work is still done,
but not by employed workers). In this respect one might argue that outsourcing tasks to countries with low labor costs does not change occupational safety in global perspective; it only reallocates hazards geographically. Thus, as long as the destructed jobs are associated with hazardous work that has become redundant, the job-pool composition-effect will represent a real change in occupational safety. If destruction of jobs is associated with a reallocation of hazardous tasks (e.g. from employed workers to self-employed workers), the job-pool composition-effect is in essence no more than a spurious change in occupational safety.

On the supply side of the market the matched worker population is equipped with hazard handling skills determining how much of the hazard potential is transformed into observed accident frequencies. Clearly, the amount of these skills present at any moment in time is subject to changes. One interesting hypothesis in this respect is for instance the effect of an ageing worker population on the accident frequencies. Presumably, as workers get older their hazard handling skills seem to deteriorate. The Dutch data show a steadily with age rising accident frequency for workers older than 30. In theory, this suggests that ageing of the worker population induces a higher accident frequency.

3 DATA

3.1 Registered population at risk

The population at risk is defined as all employed workers registered to be living in The Netherlands. Self-employed individuals and foreigners not formally registered are excluded from the analysis. Micro register data on the population at risk was supplied by Statistics Netherlands (CBS). Information of several registrations is combined in the so-called SSB, which is the Dutch abbreviation of ‘social statistical micro data’. The SSB is in essence based on the municipal registrations of inhabitants, registrations of social insurances of employed workers and income taxes. The following information was used:

- date of birth; gender; first and second generation foreign descent;
- beginning and end dates of jobs; contract hours; type of contract (flex workers)
- sector of industry (distinguishing 262 sectors).

Individuals are uniquely identified by a personal code derived from the citizen service number (BSN). Sectors of industry are identified using the Dutch Standaard Bedrijfsindeling (SBI 2008) which is based on the activity classification of the European Union (Nomenclature statistique des activités économiques dans la Communauté Européenne, NACE) and on the classification of the United Nations (International Standard Industrial Classification of All Economic Activities, ISIC). In the present paper the first three digits allow differentiating between 262 sectors of industry.

3.2 Registered accident casualties

The Dutch labor inspectorate (I-SZW) registers the casualties of serious occupational accidents. These accidents are serious in the sense that they lead to death, permanent injury or hospitalization of the casualty. Traffic accidents on the road, in the air and on sea or waterways are not included in this registration. Furthermore accidents with dangerous substances (e.g. fireworks, asbestos, radioactive materials) and natural resources (gas) are excluded. All these excluded accidents are investigated by other specialized inspectorates. The I-SZW registration however covers the vast majority of serious occupational accidents in The Netherlands. From all registered casualties in 1999-2011 only the employed workers were selected. Self-employed individuals and collateral casualties (e.g. passers-by) were excluded from the analyses.

Although reporting of accidents is mandatory/compulsory, there is suspicion of underreporting. This may occur when a serious accident is erroneously not reported to the inspectorate. The amount of underreporting is estimated to range from 1000 to 1500 accidents per year (Schouten et al, 2008). This amounts to say 30-40% of the yearly average of serious accidents investigated by I-SZW or say 25% of the ‘true’ number of accidents. However, the estimate by Schouten et al should be treated with caution, since it is based on a comparison with registered casualties of a small sample of first-aid centers unevenly spread across the country.

Since there is no evidence of a selection bias due to underreporting, we will treat the registered casualties as a representative 75%-sample of the true casualty population.

3.3 One-to-one merging

Data was assembled by one-to-one merging of several registers by means of a unique identifier of individuals: the citizen service number (BSN). The BSN-code of the casualties was not registered by I-SZW but derived by CBS on the basis of date of birth, gender and address (zip-code and house number). In 1999-2011 a total of 30,424 casualties were registered in all inspected serious accidents in The Netherlands. The citizen service number of 6,353 (20.9%) of these casualties could not be identified, mostly due to lacking or incorrect personal information. Also casualties who were not formally registered as inhabitants or as tax-paying workers could not be identified. The remaining 24,071 (=30,424–
6,353) identified casualties compare very well to the total number of casualties counted in the Storybuilder database, in which all serious accidents are bow tie modelled by hand (Bellamy et al., 2006, 2007, 2008). This suggests that the quality of gathering of personal information by I-SZW may fall short when it turns out during the course of the investigation that the incident does not formally qualify as a serious accident. In general, identification failure does not give rise to a selection bias, with the exception of lethal accidents. Since the count of lethal accidents is well-established in Storybuilder we estimate a loss of say 10 out of the expected number of 75 deaths per year. From the 24,071 identified casualties another 2,842 were excluded because they were not identified as members of the population at risk (employed workers). The final sample of occupational accident casualties consists of 21,229 individuals.

4 DEFINITIONS AND EMPIRICAL MODEL

4.1 Definitions

The time axis is measured in months, resulting in 156 time periods $t$ in the years 1999-2011. In each $t$ the actual number of workers – on average 7.5 million workers per period – is selected using the dates marking the beginning and ending of a job. The duration of exposure to occupational hazards per month is measured in full time (8 hour) working days according to the contract, denoted as $d_{it}$. For part time workers the working hours are transformed to full time working day equivalents. For instance, a half-time job of four days per week adds up to two full time working days. Workers with flexible hours are separated from other workers. Exposure is corrected for weekend days, varying length of the calendar months and Christian and national holidays. Sick leave and vacations could not be corrected for. The job hours of workers with two (or more) jobs at the same time are added up. Should these jobs be in separate sectors of industry, the sector of the largest job is assigned to the smaller ones. In each period $t$ an indicator $C_t$ expresses the state of the Dutch business cycle (DNB, 2013). A discrete indicator $A_{it}$ expressing whether member $i$ of the population at risk was registered as an accident casualty in period $t$ was constructed by one-to-one merging the casualties to the population at risk. Indicator $A_{it}$ is defined as:

$$A_{it} = \begin{cases} 
1 & \text{member } i \text{ of the population identified by data merging as casualty in } t \\
0 & \text{otherwise}
\end{cases}$$ (1)

With on average 7.5 million workers per period we observe approximately one billion $A_{it}$ in 1999-2011. In order to bring down this number to manageable size, groups of population members are defined. Each group $g$ contains all population members $i$ in period $t$ who are homogenous with respect to characteristics vector $X_i$. This vector includes the variables: age, gender, native versus non-native workers, flex versus normal workers, part time workers, new recruits and 262 sectors of industry. This results in on average 90,000 homogeneous groups $g$ in any period $t$. Note that regression-wise no information is lost since there are no variables left to differentiate within the groups. Within groups $g$ the total number of accidents $N_{gt}$ in period $t$ is defined as:

$$N_{gt} = \sum_{i\in g} A_{it}$$ (2)

and the group’s total amount of exposure to hazards measured in full time working days is calculated as:

$$D_{gt} = \sum_{i\in g} d_{it}.$$ (3)

Since $N_{gt}$ is a nonnegative integer count it can be regressed on explanatory variables by means of a count data model.

4.2 Empirical model

The literature offers a wide variety of count data models to regress a count variable on a set of independent variables (see for example Cameron and Trivedi, 1998). A well-known example is the Poisson model. The applicability of this model however is limited due to the fact that it imposes equi-dispersion, which means that the conditional mean and variance are equal. This is a rather strong assumption. A more flexible count data model is obtained when unobserved heterogeneity is introduced in the Poisson intensity parameter $\lambda$. The most commonly used extension to the Poisson model is the negative binomial model (negbin), which results when $\lambda$ is mixed with a gamma distribution. The negbin-model imposes over-dispersion, meaning that the conditional variance exceeds the mean. The model contains the Poisson model as a directly testable special case. The negbin-model is written as:

$$P(N_{gt} = y) = \frac{\Gamma(\rho^{-1} + y)}{\Gamma(\rho^{-1})\Gamma(y)} \left( \frac{\rho \lambda_{gt}}{1 + \rho \lambda_{gt}} \right)^y \left( 1 + \rho \lambda_{gt} \right)^{-\rho^{-1}}$$ (4)

where the non-negative parameter $\rho$ captures the degree of over-dispersion. If $\rho$ converges to zero, the model reduces to the Poisson model. We will use the negbin type I model, which is parameterized such that $\rho$ is scaled by $\lambda_{gt}$. This means that $\lambda$ varies across individuals and that – conditional on covariates – the count variance is a linear function of the count mean. The model is estimated using the nbreg procedure in STATA (version 12).
In the present application of the model the intensity parameter $\lambda$ is be interpreted as the mean rate of occurrence of serious occupational accidents per 8-hour working day. In short we will refer to $\lambda$ as the hazard rate of getting involved in an occupational accident. Hazard rate $\lambda$ is modelled to vary with group characteristics $X$ and may vary in time $t$. Time variation may be due to structural tendencies (as a result from changes in the matching of supply and demand on the labor market), seasonal patterns and the business cycle $C_t$. In short:

$$
\lambda_{gt} = \exp(X_{gt}, f(t), C_t, D_{gt})
$$

(5)

with the log of $D_{gt}$ enclosed as an offset to correct for group differences in exposure and $f(t)$ capturing a yearly trend and a seasonal pattern. A seasonal pattern is modelled by dummy variables for the twelve months. However these estimates are likely to be biased by the fact that season-biased sick leave and vacations cannot be corrected for in $D_{gt}$. For instance, since many workers spend most of their time off during the summer holidays, exposure $D_{gt}$ is probably overestimated in July and August leading to a (spurious) decline in the hazard rate $\lambda$ in these months. Therefore, any observed seasonal pattern in $\lambda$ is attributed to unobserved patterns in absence from work due to sickness and vacations.

In the specification of eq. (5) mainly 0/1-indicator variables are included. Initially, indicator variables for all ages ranging from 15 to 64 were estimated, revealing a declining accident frequency only for worker between 20 and 30 years of age. Therefore, in the final model linear effects were estimated for three age intervals: 15-19, 20-29 and 30-64. The effect of job tenure is estimated discretely using an indicator for ‘new recruits’. New recruits are defined as workers not registered to have been working in the preceding year. In each age interval a fixed effect for new recruits is estimated. Three job size intervals are specified: part-time workers of 1-49% and 50-89% of a working week and full-timers (90-100%). The model also includes 262 coefficients for sectors of industry, one for male workers, one for flex workers, one for non-native workers and eleven for the months February to December (January was chosen as reference). A linear time trend is included by counting the years as of 1999.

4.3 Hazard ratio

We will denote the estimated coefficients by $\beta$. Whenever $\beta$ refers to a 0/1-indicator variable $\exp(\beta)$ can be interpreted as a hazard ratio. For instance, let $\beta$ be estimated coefficient for male workers. Hazard ratio $\exp(\beta)$ then expresses the relation between the incidence rate of male and female workers. An estimate of say 0.69 would imply that the incidence rate of occupational accidents among male workers is twice the incidence rate of females since $\exp(0.69)=2$. We will express the hazard ratio with respect to a reference group (within the indicator set) with the mean hazard rate closest to the population average.

5 ESTIMATION RESULTS

In this section the results of the estimated negbin-model are discussed. The estimated model results in 286 coefficients of which 262 are fixed-effects for sectors of industry. We will address these sector effects separately by means of descriptive statistics in subsection 5.3. In subsections 5.1 and 5.2 we will discuss the remaining coefficients reflecting differences in age, gender and nativity and the worker-firm relationship. Subsection 5.4 focusses on the question to what extent the results can be interpreted as an achieved improvement in Dutch occupational safety.

5.1 Worker heterogeneity

The average hazard rate of serious occupational accidents in The Netherlands in 1999-2011 per 8-hour working day is estimated at $1.03 \times 10^{-6}$. Taking into account a possible underreporting of 30-40 percent the mean hazard rate would be approximately $1.4 \times 10^{-6}$. That means 1.4 serious accidents on one million full time working days. The hazard rate varies strongly over the population at risk. The distribution of hazard rates is skewed to the right, showing a relatively large amount of mass beneath the mean. The median is less than half the mean: $0.6 \times 10^{-6}$.

The ratio of the hazard rates defining the 1st and 99th percentile of the worker population is approximately 700. This means that the risk determinants incorporated in the model (age, sex, nativity, worker-firm relationship and sector) can account for large risk differentials within the worker population. However, further improvement of the model could be achieved if workers within the sectors are differentiated according to the hazard potential of their jobs, for instance by means of the International Standard Classification of Occupations (ISCO). This information was not available in our data.

The male-female hazard ratio is estimated at 4.2, implying male hazard rates of on average say four times the female hazard rate. It is likely that this large hazard ratio is to some extent due to selection of male workers in jobs with high hazard potential (in for instance construction work and industry). A remarkable hazard-age pattern is observed. An initial specification of the model with separate fixed-effects for all ages revealed that the hazard rate peaks at the age of 19 and subsequently declines...
with age in the interval 20-29. For workers older than 30 the hazards rate increases steadily as they grow older. On the basis of this observation separate linear effects were estimated for the intervals 15-19, 20-29 and 30-64. Results are displayed in Table 1.

Table 1. Estimated age effects

<table>
<thead>
<tr>
<th>age 15-19</th>
<th>age 20-29</th>
<th>age 30-64</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.088</td>
<td>-0.023</td>
<td>0.006</td>
</tr>
<tr>
<td>0.028</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>1.09</td>
<td>0.98</td>
<td>1.01</td>
</tr>
<tr>
<td>1.42</td>
<td>0.81</td>
<td>1.23</td>
</tr>
</tbody>
</table>

As Table 1 shows, the hazard rate increases at a rate of 9 percent per year for workers younger than 20. The hazard ratio for workers of 15 and 19 years of age equals 1.42, implying a 42 percent higher rate for the 19-year old workers compared to workers 15 years of age. An explanation for this remarkable difference may be found in the fact that young students in low vocation education are formally employed by firms but spend half of their contractual working hours in school. As their education evolves, their actually worked hours converges to full time hours. This suppresses their true exposure to hazards at younger ages, resulting in low estimated hazard rates. Since trainees and their true working hours cannot be identified in our data, this explanation could not be elaborated on in more detail. Presumably, effects of dual vocational education is minimal for young workers older than 20, since most programs end before 20.

For workers 20-64 year of age the hazard-age pattern is U-shaped. In the first ten years the hazard shows a 19% decline of the hazard from age 20 to 29. Subsequently the hazard gradually rises as workers grow older. From age 30 to 64 an increase of 23 percent is estimated. The estimated effects are significant beyond statistical doubt. The average hazard rate for all age cohorts is depicted in Figure 1.

The hazard ratio of non-native workers in The Netherlands equals 1.3, implying a significantly higher accident rate among workers with at least one parent born in a foreign country. In a preliminary specification of the model (not presented in this paper) an effect was tested for second generation non-natives. For those workers the hazard rate turned out to be approximately equal to the hazard rate of Dutch natives. From this one may conjecture that later generations of non-native workers:

- end up in less hazardous jobs than earlier generations, and/or
- are better equipped with hazard handling skills.

The literature provides little solid empirical evidence supporting relatively high hazard rates of foreign/migrant workers. In Berkhout et al. (2014) an analysis of foreign accident casualties in the Netherlands shows that 15% of all casualties was foreign, whereas their share in the total number of worked days is estimated at 5-10%, depending on how one deals with multiple nationalities. They show that most of the differential can be accounted for by selection of foreigners in high-risk sectors. High accident rates of foreigners prevail in some high-risk sectors, such as the construction industry. In Dressler (2012) it is argued that there is a relation between occupational risk and the extent to which workers are familiar with safety regulations, cultural settings and the dominant language at the workplace. In the present framework migrant workers are in that case on average less endowed with hazard handling skills. On the other hand, it is sometimes claimed that cultural differences will force foreign workers to work more safely than their local counterparts, grounded by a sense of responsibility towards the family and their financial needs (Gron, 2012).

5.2 Worker-firm relationship

Whether the worker-firm relationship is associated with hazard rate differentials is tested by three variables: job size (i.e. hours worked per week), flexible hours and tenure. Estimates indicate that ceteris paribus the hazard rate tends to decrease with increasing job size. That is, part time workers have higher hazard rates than full timers. See Figure 2.
The hazard ratio of flex-workers is estimated at 1.3, indicating that workers flexibly affiliated to the firm tend to have higher occupational risk. A job tenure effect was tested by separating new recruits – those entering the labor force for the first time or after an absence of at least one year – from the more experienced workers. Effects in the three age intervals 15-19, 20-29 and 30-64 were estimated, none of which were found significant by conventional standards. This means we find no evidence of higher occupational risk for new recruits in general.

5.3 Sector hazard ratios

Sectors of industry were defined using the first three digits of International Standard Industrial Classification of All Economic Activities (ISIC). This results in 262 sectors for all of which a fixed-effect is estimated. For all sectors the hazard ratio with respect to sector Manufacture of Communication Equipment (ISIC-code 26.3) is calculated. This sector is chosen as reference since its average hazard rate is closest to the total weighted population average.

By definition the sector hazard ratios express average job hazard potential differentials, controlled for worker selection. However, sector differentials will in fact be larger if male workers are selected into hazardous jobs. That is, if jobs with high hazard potential are mostly done by male workers. If that would be the case, the estimated male-female hazard ratio merely reflects this selection effect rather than a gender difference in hazard handling skills. A high percentage of male workers in a sector then presumably signals the presence of hazardous jobs. Therefore an alternative sector hazard ratio incorporating sector differences in male-female composition is presented. This correction inflates the standard deviation of the sector hazard ratios from 1.4 to 1.7. The maximum hazard ratio increases from 6.5 to 7.9. Descriptive statistics of the 262 gender-corrected sector hazard ratios are displayed in Table 2.

By definition the sector hazard ratios express average job hazard potential differentials, controlled for worker selection. However, sector differentials will in fact be larger if male workers are selected into hazardous jobs. That is, if jobs with high hazard potential are mostly done by male workers. If that would be the case, the estimated male-female hazard ratio merely reflects this selection effect rather than a gender difference in hazard handling skills. A high percentage of male workers in a sector then presumably signals the presence of hazardous jobs. Therefore an alternative sector hazard ratio incorporating sector differences in male-female composition is presented. This correction inflates the standard deviation of the sector hazard ratios from 1.4 to 1.7. The maximum hazard ratio increases from 6.5 to 7.9. Descriptive statistics of the 262 gender-corrected sector hazard ratios are displayed in Table 2.

The sectors Manufacture of Weapons & Ammunition (25.4) and Other First Processing of Steel (24.3) show the highest hazard ratio and can be regarded as the most risky sectors to work in. The top 10 high-hazard sectors features three branches of the metals industry, two mining branches and two branches closely related to construction: Demolition & site preparation (43.1) and Manufacture of articles of concrete, cement and plaster (23.6). Various branches in the construction sector show hazard ratios ranging from 4.0 to 5.2.

<table>
<thead>
<tr>
<th>Sector hazard ratios (male-female composition included)</th>
</tr>
</thead>
<tbody>
<tr>
<td>average</td>
</tr>
<tr>
<td>standard deviation</td>
</tr>
<tr>
<td>minimum</td>
</tr>
<tr>
<td>maximum</td>
</tr>
<tr>
<td>N (sectors)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

10 highest hazard ratios w.r.t. Manufacture of Communication Equipment (ISIC 26.3)

Manufacture of weapons and ammunition (25.4) 7.9
Other first processing of steel (24.3) 7.8
Casting of metal (24.5) 6.9
Manufacture of basic precious and other non-ferrous metals (24.4) 6.4
Mining of stone, sand and clay (08.1) 6.2
Demolition and site preparation (43.1) 6.1
Materials recovery (38.3) 6.1
Manufacture of articles of concrete, cement and plaster (23.6) 6.0
Sawmilling and planing of wood (16.1) 6.0
Other mining and quarrying (08.9) 6.0

Occupational accidents are unevenly distributed over the population at risk. This can be shown graphically by means of a Lorentz curve, which is commonly used in economics to analyze inequality in the distribution of wealth and income. Inequality can be measured by the Gini-coefficient, calculated as the ratio of the area enclosed by the curve and diagonal and the total area beneath the diagonal. For the sector distribution of occupational accidents in The Netherlands the Gini-coefficient is estimated at 0.37. In Figure 3 the Lorentz curve of occupational accidents is depicted.
For example, from Figure 3 we have inferred that 36 percent of all serious occupational accidents are registered in just a handful of sectors representing 14 percent of total employment. These sectors (in the upper right corner of Figure 3) are:

<table>
<thead>
<tr>
<th>Sector (ISIC-code)</th>
<th>Share in accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporary employment agencies job pools (78.2)</td>
<td>8.8%</td>
</tr>
<tr>
<td>Roofing and other specialized construction (43.9)</td>
<td>4.6%</td>
</tr>
<tr>
<td>Construction installation (43.2)</td>
<td>4.0%</td>
</tr>
<tr>
<td>Construction of buildings (41.2);</td>
<td>3.6%</td>
</tr>
<tr>
<td>Manufacture of structural metal products (25.1);</td>
<td>2.7%</td>
</tr>
<tr>
<td>Other specialized wholesale (46.7);</td>
<td>2.6%</td>
</tr>
<tr>
<td>Building completion (43.3);</td>
<td>2.5%</td>
</tr>
<tr>
<td>Construction of roads, railways, bridges (42.1);</td>
<td>2.4%</td>
</tr>
<tr>
<td>Freight transport by road (49.4);</td>
<td>2.4%</td>
</tr>
<tr>
<td>Cleaning activities (81.2)</td>
<td>2.2%</td>
</tr>
</tbody>
</table>

Temporary employment agencies are at the top of the list. This implies that temporary workers can be considered a high-risk group of workers that may need special attention in preventive policy measures.

5.4 Effects of structural changes in labor supply and demand and the business-cycle

Over the years 1999-2011 a significant decline of the hazard rate of 1.9% per year is estimated. This adds up to a total decline of 27% between 1999 and 2011. The observed population average hazard rate decreased from 1.60 x 10^-6 in 1999 to 1.20 x 10^-6 in 2011, including structural changes in labor supply and demand. On the supply side ageing of labor supply affects the hazard rate if hazard handling skills of workers gradually deteriorate when they grow older. Between 1999 and 2011 the average age of workers increased from 36.3 to 39.5. The share of workers older than 30 increased from 67.4% to 72.4%. As was mentioned in section 5.1 (see Table 1), the hazard rate of workers older than 30 significantly increases each year by 0.6%. From this we estimate an ageing effect of approximately 2%. Thus, without ageing the average hazard rate in 2011 would have been approximately 1.15 x 10^-6 instead of the observed 1.20 x 10^-6. On the other side of the market structural changes in the sector composition of labor demand affect the population average hazard rate. Holding the sector composition of 1999 constant, the hazard rate corrected for ageing is estimated at 1.30 x 10^-6 in 2011.

We conclude that in the absence structural changes in labor demand and supply the population average hazard rate would have decreased by 19% in 13 years. This may be interpreted as the achieved improvement of occupational safety in 1999-2011. The observed 27% decline between 1999 and 2011 is obscured by a small ageing bias (+2%) on the one hand and a somewhat larger hazard-biased destruction of jobs on the other hand (-10%). The former pushes total decline downward (towards zero), since older workers are associated with high hazard rates. The latter amplifies the decline, as jobs with high hazard potential are more often subject to job destruction.

A business cycle effect was tested by incorporating a monthly national business cycle indicator in the model. The results show a positive significant effect. That is, in an economic boom the accident frequency is 4-6% higher than in a slump.

6 CONCLUDING REMARKS

This paper sets out to quantify the occupational risk of all registered employed workers, based on merged registration data. By means of a negative binomial count data model cross-sectional differences as well as time trends are analyzed. In a cross-sectional perspective the model predicts a hazard ratio between the 1st and 99th percentile of the worker population of approximately 700 on the basis of age, sex, nativity, worker-firm relationship and sector. A 27% decline in the hazard rate over the years is partly explained by structural changes in labor supply and demand (e.g. ageing and hazard-biased jobs destruction). All in all, a considerable 'unexplained' improvement of occupational safety in The Netherlands remains.

In The Netherlands the use of registration data for statistical purposes has become common practice in the past decade. In this study a registration of casualties of serious occupational accidents was merged to the entire population of employed workers. The analysis combines data from individuals, jobs, firms and casualties. As the paper shows, the registration data can be transformed into accurate policy information when the proper statistical tools are utilized. Accuracy of the results would improve considerably if a job classification could be made available. But all in all, the paper shows that the hazard rate of individual workers can be assessed effectively by combining registration data. We consider this the main contribution of the paper to the literature.

Finally, we have shown that occupational accidents may be analyzed within the standard labor economics framework of matching skills (human capital) to job requirements. The skills needed to handle a job’s hazard potential – say the hazard handling skills – may be treated as any productive - revenue generating skills workers are endowed with. Since it is practically impossible to measure the hazard potential of jobs, we believe the focus of attention should be on the actual hazard rate which results from particular job-worker matches. The paper shows that by using complete registrations we are able to identify the effects of structural changes in labor supply and demand on occupational safety in national perspective.
REFERENCES


DNB (2013), DNB-business cycle indicator, ww.dnb.nl

Dressler, D. 2012. Increased occupational risks among migrant workers in Austria, ISSA, WOS 2012

Grön, S., and G. Cruz 2012. National differences in safety practice, Centre for Maritime Health and Safety, University of Southern Denmark, University of the Phillipines. WOS 2012

EU-O SHA (2014)


Schouten, M.J., A. Faas en J. Hoeben (2008), Achtergronden dodelijke en ernstige arbeidsongevallen op basis van in 2007 afgesloten ongevalsonderzoeken, Arbeidsinspectie, december 2008 (in Dutch)


United Nations (2008), International Standard Industrial Classification of All Economic Activities, Statistical Papers, Series M, revision 4