SAFETY SCIENCE Monitor

Issue 2 2015

Article 4

VOL 19

DESIGNING AN INDIVIDUAL RISK ASSESSMENT TOOL BASED ON REGISTRATION DATA

P.H.G. BERKHOUT
RIGO Policy research and consultancy, Amsterdam, The Netherlands

J. BUITENDIJK
RIGO Policy research and consultancy, Amsterdam, The Netherlands

M. DAMEN
RIGO Policy research and consultancy, Amsterdam, The Netherlands

ABSTRACT

In this paper a design for an individual risk assessment tool is proposed. The occupational risk of individual employed workers is empirically quantified using merged registration data covering the entire population at risk. The analysis features a negative binomial count data regression model for statistical inference. In the model the mean rate of accident occurrence per working day is determined by variables characterising the personal background and working conditions, such as age, gender, contract type, working hours and sector. The estimated model is used to build a computer tool that can assist employers in assessing individual occupational risk. Since serious occupational accidents are rare events from the perspective of one single employer, he is not in the position to accurately assess the occupational risk of an individual worker on the basis of his own experience. For that he needs accident information drawn from the entire population enabling him to reflect on his occupational safety compared to others. Therefore the tool features two risk indicators: one positioning the individual's risk in the entire working population of employed workers and one positioning the individual in the branch of industry in question. This way the employer is made aware of the fact that given the occupational risk level in his branch, there may be room for improvement of occupational safety by optimizing job-worker matches.

An important contribution of the paper to the literature is the fact that calculations are based on registration data only and include the entire population at risk, comprising of all employed workers registered to be living in the Netherlands in 1999-2011. Data were assembled by means of one-to-one match merging of complete registration data on individuals, jobs, firms and accident casualties. The research was initiated by The Dutch National Institute of Public Health and the Environment (RIVM) in order to identify long term trends in serious occupational accidents in The Netherlands.

On average, the mean hazard rate of serious occupational accidents per working day is estimated at approximately 1.4×10^{-6}. This refers to accidents leading to permanent injury, hospitalisation or death of the worker. Hazard rates vary strongly between workers. Risk is found to be significantly related to age, gender, nativity, tenure, contract type, working hours and sector. The explanatory variables in the model predict a hazard ratio of 700 between the 1st-percentile and 99th-percentile of the worker population. Analysis of high-risk branches shows that 36% of all accidents are associated with 10 branches (out of 262) accounting for only 14% of employment.

Key words: occupational risk assessment, serious occupational accidents, econometric modelling, hazard rate, registration data, high-risk groups, risk tool, hazard handling skills

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. However, in practice employers do not have the quantitative means to assess the occupational risk of individual workers other than their own experience. Since serious occupational accidents are rare events from the perspective of one single employer, it is unlikely that employers are able to accurately assess the occupational risk of an individual worker. An empirically based computer tool may therefore provide employers with the necessary information to reflect on their position in the risk distribution.

In the present paper the individual rate of accident occurrence per working day is estimated by merging complete registration data on casualties of serious occupational accidents to individual registration data describing the entire population at risk. Individual occupational risk is assessed by means of a negative binomial count data model. In this regression 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 and the male/female ratio of the job).

Since the data cover a time span of 13 years the estimation results can be used to analyse changes in Dutch occupational safety in time. Result are presented in Berkhout and Damen (2014). In a nutshell the conclusion are the following. A remarkable decline of 27% is observed in the average rate of accident occurrence, raising the question whether occupational safety in The Netherlands has truly improved to such extent. Within the conceptual framework of labour market matching of skilled workers to job requirements, it is conjectured that structural changes in labour 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 not negligible, due to a serious 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 favour of the less hazardous jobs. Contrary to the ageing-effect a lower average rate of accident occurrence results. Roughly estimated this composition-effect accounts for one third the total decline. All in all, an unexplained decline in the accident rate of 19% in 13 years remains.

In the present paper we will demonstrate that the estimation results can be used to build an individual risk assessment computer tool enabling employers to calculate the individual hazard rate and compare this rate to a relevant reference population. In other words, the tool may help to pinpoint the relative location of any individual worker in the risk distribution.

The paper is structured as follows. In section 2 the observed individual hazard rate is considered in the conceptual framework of labour market matching of workers to jobs. Data, definitions and a formal description of our empirical model are presented in sections 3 and 4. Section 5 presents a brief overview of the statistical results. In section 6 the risk assessment tool is presented. Section 7 concludes.

2. CONCEPTUAL FRAMEWORK: LABOUR MARKET MATCHING OF SKILLS TO JOB REQUIREMENTS

The idea of a quantitative assessment tool for employers presumes that they may contribute to occupational safety once they are provided with the proper information. A conceptual framework is needed to interpret the results of a statistical model. In this paper the rate of occurrence of occupation-al accidents – the accident frequency– is analysed within the human capital framework of labour market matching of workers to jobs. On the labour 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.

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 branches of industry).

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 branches of industry.

3.2. Registered accident casualties

The Dutch labour 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 centres 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 match 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 timeperiods 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 hours) working days according to the contract, denoted as d_{it}. For part time workers the working hours are transformed to full time equivalents of a working day. 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

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

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 (measured in full years on the first day of month t), gender, native versus non-native workers, flex versus normal workers, part time versus full 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:

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

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

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

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 such a 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 λ. The most commonly used extension to the Poisson model is the negative binomial model (negbin), which results when λ 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

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

where the non-negative parameter ρ captures the degree of over-dispersion. If ρ converges to zero, the model reduces to the Poisson model. We will use the negbin type I model, which is parameterised such that ρ is scaled by λ_{gt}. This means that ρ 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 λ is be interpreted as the mean rate of occurrence of serious occupational accidents per 8-hour working day. We may also refer to λ as the hazard rate. Hazard rate λ 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 labour market), seasonal patterns and the business cycle C_t.

$$\lambda_{gt} = \exp(X_g, f(t), C_t, D_{gt}) \quad (5)$$

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 accident frequency λ in these months. Therefore, any observed seasonal pattern in λ 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.

Estimated coefficients are denoted by β. Whenever β refers to a 0/1-indicator variable, the *hazard ratio* is equal to $\exp(\beta)$. For instance, let β be the 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 $\exp(0.69)=2$.

4.3 Calculating relative risk indicators

In the present paper the estimated regression coefficients β are used to develop a relative risk assessment tool. In this tool relative risk is expressed in terms of two indicators: 1) one representing the fraction of workers in the worker's own sector of industry with higher hazard rate; and 2) one indicator representing the fraction of workers in the entire worker population with a higher hazard rate. The tool should be helpful in the process of assessing individual risk as it locates the position of a worker within two relevant risk distributions, depending on his characteristics X. Let P_s and P_{total} denote the worker's position in the risk distributions of sector s and the total worker population respectively. Supposing a log-normally distributed hazard rate P_s and P_{total} are defined as

$$P_k = 1 - \Phi\left(\frac{X'\beta - \mu_k}{\sigma_k}\right) \qquad k = s, total \qquad (6)$$

where μ en σ are the sector-specific parameters of the corresponding normal distributions.

5. A BRIEF OVERVIEW OF MODEL ESTIMATION RESULTS

Results of the model estimation are described in detail in Berkhout and Damen (2014). In this section a brief overview of the results are discussed. The average hazard rate of serious occupational accidents in The Netherlands in 1999-2011 per 8-hour working day is estimated at 1.0 x 10^{-6}. Taking into account a possible underreporting of 30-40 percent the mean hazard rate would be approximately 1.4 x 10^{-6}. That means 1.4 serious accidents on every 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 x 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, non-native workers, flex workers, new recruits, job size and sector) can account for large risk differentials within the worker population.

The male-female hazard ratio is estimated at 4.2, implying male hazard rates of on average roughly four times the female hazard rate. It is likely that this rather 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). 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. A remarkable hazard-age pattern is observed. 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. The average hazard rate for all age cohorts are depicted in Figure 1.

Figure 1. The average hazard rate and worker age.

Furthermore, estimates indicate that the hazard rate tends to decrease with increasing job size. That is, part time workers have higher hazard rates than full timers. The hazard ratio of half time (compared to full time workers, ceteris paribus) is approximately 1.4, implying a 40 percent higher hazard rate for half time workers compared to full time workers. The hazard ratio of flex-workers is estimated at 1.3, indicating that workers flexibly affiliated to the employer's firm tend to have higher occupational risk. A new recruit's effect was tested by separating workers entering the labour force for the first time or after an absence of at least one year from the more experienced workers. No evidence of higher occupational risk for new recruits was found.

Finally, the model estimations results can assist in identifying high-risk branches and firms. It turns out that 36% of all accidents are associated with 10 branches (out of 262) accounting for only 14% of employment.

6. AN APPLICATION: A RELATIVE RISK ASSESSMENT TOOL

The process of risk assessment commences in the first step with the identification of hazards and the workers at risk. However, as suggested in the framework (section 2) this assessment is in the present application not limited to the hazard potential intrinsic to the job. In our view the hazard rate is associated with a particular

worker-job match. Therefore, the assessment also takes into account personal characteristics of the worker, instrumenting his human capital concerning hazard handling skills. In the proposed relative risk assessment tool the user chooses one out 262 sectors of industry. Since scrolling through 262 sectors is rather tedious, the user first chooses a main sector (out of 24) and subsequently a specific branch is chosen within the main sector. Subsequently the user fills in the worker's age, nativity, type of contract (flexible hours) and the number of hours worked per week. The tool does not enable the user to differentiate between various types of jobs within his firm. Presuming a selection of male workers in jobs with high hazard potential, the hazardous nature of the job is characterised by tuning the male-female ratio. That is, a high male-female ratio is associated with high-risk jobs, and vice versa. Figure 2 shows the input (upper panel) and output (bottom panel) of the proposed relative risk assessment tool.

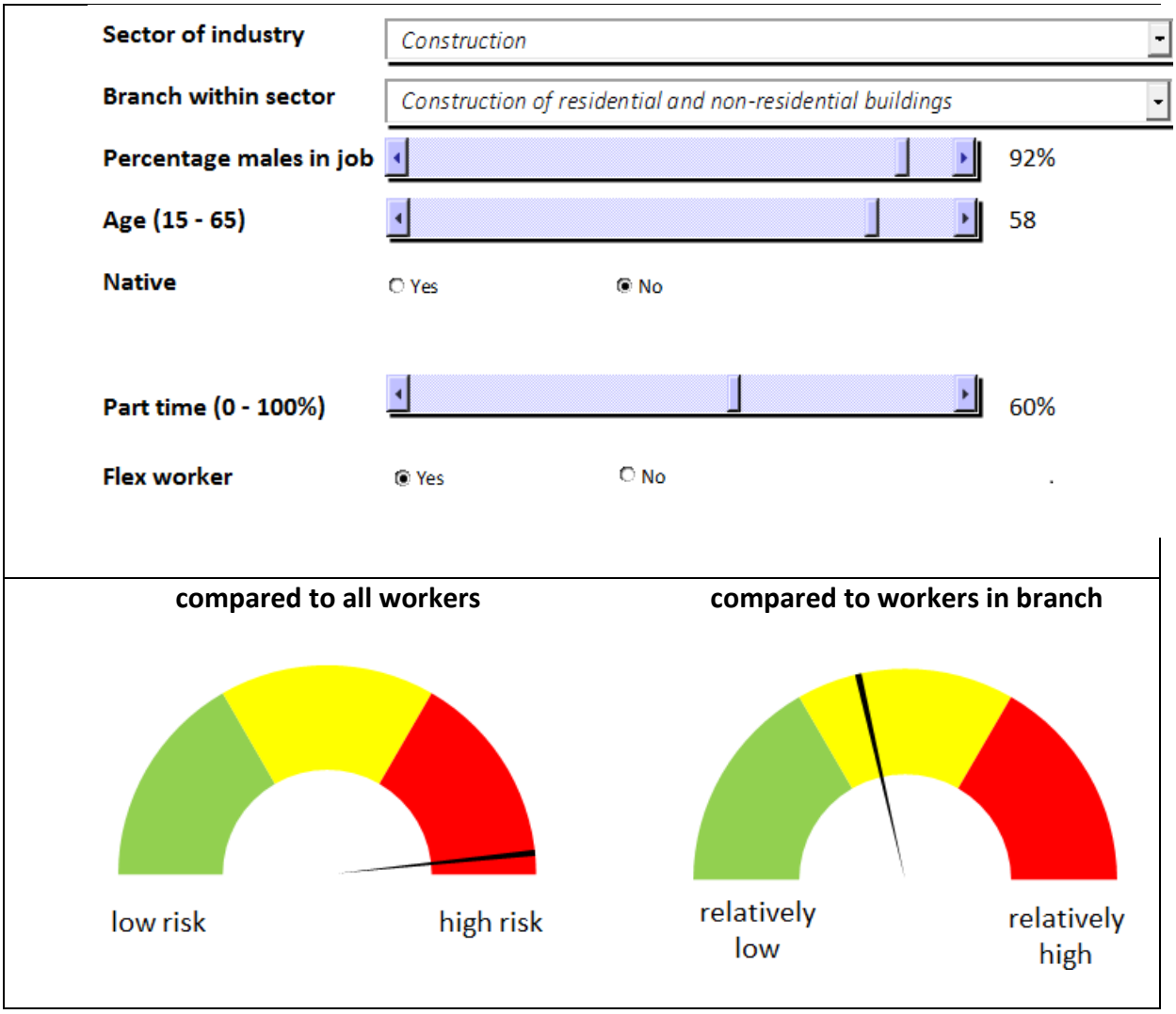

Figure 2 Designing a relative risk assessment tool: input (top) and output (bottom)

7. CONCLUDING REMARKS

Risk assessment is a cornerstone of the European approach to prevent occupational accidents and ill health. However, in practice employers do not have the quantitative means to assess the occupational risk of individual workers other than their own experience. Since serious occupational accidents are rare events from the perspective of one single employer, it is unlikely that employers are able to accurately assess the occupational risk of an individual worker. This paper demonstrates that a simple computer tool, based on national registration data only, can effectively assist the employer to this important task. The tool makes the user aware of the position of the individual in the entire national risk distribution as well as the relative position in the branch of industry. A clear advantage of this dual approach is that also in relatively safe branches the tool may signal potential risk.

The here presented prototype requires the input of six variables: branch of industry, job type, age of the worker, nativity, number of hours per week and the worker-firm relationship (flex worker or not). Since a classification of jobs was not available in our data, jobs are characterized by the percentage of males in the job (in

deviation from the branch mean). This was based on the observation that hazard rates of males are on average four times the rates of females. We conjecture this to be mainly the result of selection of male workers into high-risk jobs. A high percentage of males in some type of job then signals high occupational risk. It goes without saying that inclusion of a discrete job classification in our model will improve the tools practical value.

REFERENCES

Bellamy L.J., Oh J.I.H., Ale B.J.M., Whiston J.Y., Mud M.L., Baksteen H., Hale A.R., Papazoglou I.A , (2006), Storybuilder: The New Interface for Accident Analysis. Proceedings of the 8th International Conference on Probabilistic Safety Assessment and Management [PSAM], May 13-19, 2006, New Orleans, ASME, New York, 2006, ISBN 0-7918-0244

Bellamy, L.J., Ale B.J.M., Geyer T.A.W., Goossens L.H.J., Hale A.R., Oh J.,Mud, M., Bloemhof A, Papazoglou I.A., Whiston J.Y. (2007), Storybuilder¬A tool for the analysis of accident reports, Reliability Engineering and System Safety 92 (2007) 735–744

Bellamy L.J., Ale B.J.M., Whiston, J.Y., Mud M.L., Baksteen H., Hale A.R., Papazoglou I.A., Bloemhoff A., Damen M. and Oh J.I.H. (2008). The software tool storybuilder and the analysis of the horrible stories of occupationalaccidents. Safety Science 46 (2008) 186-197.

Berkhout P.H.G and M. Damen (2014), Estimating individual occupational risk using registration data, Paper presented at ESREL 2014, Wroclaw Poland

Cameron, A. Colin, and Pravid K. Trivedi (1998), Regression Analysis of Count Data, Cambridge University Press, New York, 1998

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

EU-OSHA (2014) https://osha.europa.eu/en/topics/riskassessment

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