Szymon Ordysiński M.Sc.

Doctoral thesis entitled "*Identification of predictors of the severity of accidents at work, based on selected national existing data"*

Summary

The purpose of the thesis is to identify circumstances of the work being performed by the person injured at the time of the accident at work, on the basis of which the severity of the accident can be effectively predicted. To achieve this goal, in-depth analysis of accidents at work statistical data, registered in 2005-2015 by the Central Statistical Office (Główny Urząd Statystyczny, GUS), were conducted. Predicted variable was the severity of an accident at work, understood as the number of days lost due to the accident. Explanatory variables, i.e. predictors, were circumstances of the work being performed by the worker at the time of the accident, such as characteristics of the worker and working environment in which the accident at work occurred or the type of activity performed by the worker, along with the material factor connected with this activity. Verification of the relationship between the work performed by worker and the days lost due to the accident, based on the measures of the effect size, enabled identification explanatory variables that effectively allows to predict accident severity in particular groups of employees.

Critical analysis of methods and techniques of statistical analysis of accidents at work data that are in use today revealed their limitations on identification of predictors on the basis of which the severity of the accident can be effectively predicted. For this reason, in the dissertation, it was attempted to apply a new approach to accidents at work data analysis, using advanced and multidimensional methods, diversified scales and classifications, operationalized criteria for the accident severity and long periods of data recording.

As a result of the conducted analyzes, the circumstances of the work performed by the person injured at the time of the accident were identified, on the basis of which the seriousness of the accident defined by days lost can be effectively predicted. One-dimensional data analysis revealed that the most effective predictor of the severity of an accident at work is the material factor of the activity – sub-group (eta square = 0.047). Relatively well, it is also possible to predict the length of the accident absence on the basis of information on the economy sector

in which the accident occurred (eta square = 0.044). It is moderately effective to anticipate the length of absence on the basis of the profession performed by the victim – medium occupational groups (eta square = 0.038) – the accident place (eta square = 0.036) or performed work process (eta square = 0.029). On the other hand the predictive ability of the variable „the age of the victim" is small (regardless of the level of measurement). There is not a sufficiently strong relationship of the number of days lost due to the accident and remaining variables, that would allow to effectively predict the severity of accidents based on their value.

The application of multidimensional data analysis methods enabled more effective identification of predictors of the severity of accidents at work. Multidimensional methods allow to identify groups of persons injured in accidents at work, based on several variables at the same time, enabling to classify workers into groups that are more diverse in length of absence at work due to accident. The conducted analysis showed that the simple data segmentation is insufficient, and even the Twostep Cluster grouping method, which is exceptionally well adapted to the analysis of data on accidents at work, does not allow to identify work circumstances on basis of which one can accurately predict the severity of accidents at work. Much better results were achieved by combining two methods of classification trees: prior classification of predictors' values, based on CHAID trees method, which combine predictor's values that are similar in terms of days of absence at work, and then subsequent identification of persons' injured group that are differentiated in terms of days lost due to accident, based on the CRT tree method, using already transformed predictors. This method of analysis allowed to identify groups of persons injured in accidents at work, which are much more diversified due to the length of the work absence, compared to the groups defined on the basis of a single variable. The use of such a set of predictors allows you to almost twice as effectively predict the severity of accidents at work (eta square = 0.084), compared to the most effective predictor, based on a single variable.

CRT trees classification model of persons injured in accidents at work made it possible to identify the groups that are most differentiated due to the length of the work absence, however, it does not allow to determine the overall effectiveness of the predictors in forecasting the severity of all registered accidents at work. In order to determine this effectiveness, the structural equations modelling was carried out in which the relationship

between the predictors of the accidents' at work severity and the number of days lost was examined.

The structural equations model (SEM) was created on the basis of predictors, that were already transformed by CHAID classification trees. The developed SEM model presents multidimensional relations between six, previously transformed predictors, which together indicate the threat of long-term incapacity to work as a result of an accident, under given work circumstances. Calculated model allows to predict the severity of an accident at work better (standardized regression coefficient = 0.24) than the most effective single predictor. In addition, the SEM model showed that predicting the length of days lots due to the accident on the basis of several variables, that describes the circumstances of work, indicates a different effectiveness of particular predictors, than it was indicted by one-dimensional analysis. For example, circumstances of the work described by the variable "material factor of activity", which in the bivariate analysis was identified as the most effective predictor of the severity of accidents at work, in the SEM model had smaller effect on the risk of long-term inability to work. The developed SEM model showed that if the circumstances of the victim's work are described by six variables at the same time, the length of inability to work most strongly depends on the type of place of accident and economic activity of enterprise in which the accident occurred, and it depends less on the tenure, the work process and, as already mentioned, the material factor of the activity and the age of the person injured. Differences in the effectiveness of the predictors of accident severity evaluated as a result of one-dimensional and multidimensional analyzes occur because the multidimensional model takes also into an account the indirect influence of variables (relationships between individual values of predictors) and thus enables more detailed identification of aggrieved groups, that are more diversified due to the days lost due to the accidents.