

# Analysis of Occupational Accidents: A Data Mining Study

Vahideh Dadfarma<sup>1</sup>, Ahmad Soltanzadeh<sup>2\*</sup>, Samira Ghiyasi<sup>3</sup>

<sup>1</sup> Department of Health, Safety, and Environment, Faculty of Engineering, Islamic Azad University, Central Tehran Branch, Tehran, Iran • <sup>2</sup> Department of Occupational Health and Safety Engineering, Research Center for Environmental Pollutants, Faculty of Health, Qom University of Medical Sciences, Qom, Iran • <sup>3</sup> Department of Environmental Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran • \*Corresponding Author: Ahmad Soltanzadeh, Email: soltanzadeh.ahmad@gmail.com

## ABSTRACT

**Introduction:** Despite efforts exerted in various areas of the industry to reduce occupational accidents, the frequency of these accidents is reported to be catastrophically high. Therefore, this study was designed and conducted to analyze occupational accidents during a 3-year period. **Methods:** The current study was a retrospective and descriptive-analytical study carried out in four regions of Tehran between 2019-2020. The data collection instrument consisted of 818 reports of the occupational accident recorded in the Department of Labor in Tehran. Feature selection (IBM SPSS Modeler software) and binominal multiple logistic regression analysis (IBM SPSS software) were utilized in this study. **Results:** The means of age and experience of injured workers were found to be  $34.55 \pm 11.55$  and  $14.12 \pm 9.87$ , respectively. The highest rate of occupational accidents belonged to construction workshops (52.4%), production (24.4 %) and other public and social services activities (11.1 %). Data mining and modeling of factors affecting these occupational accidents showed that the consequence of the accidents was affected by seven factors ( $p < 0.05$ ). The results showed that the work experience, type of activity, number of workers, accident time on the day, type of incidence and the causes of the accident remained in the final model and were significant with the consequence of the occupational accidents ( $p < 0.05$ ). **Conclusion:** The results indicated that different parameters can affect the occurrence of occupational accidents. Additionally, the consequences of these occupational accidents can be influenced by different parameters and factors.

**Key words:** Safety; Occupational accidents; Data mining; Retrospective study

## Introduction

The incidents caused by work are among the bad consequences that have been threatened by the development of modern industries and technologies, human life and especially workers. Recognition of the risk creating factors in jobs and working environments and improving the working environment will lead to the prevention of accidents and diseases from major functions of the industry authorities as well as health personnel. In order for

governments to invest in their limited resources, they need to be aware of the imposed charges in this regard. In today's advanced world, where all matters depend on advanced and high - risk technologies, it is always feared that accidents caused by the work will cause irreparable damages. The rapid industrialization and lack of serious attention to the principles of industrial safety have led to a rise in the

**Citation:** Dadfarma V, Soltanzadeh A, Ghiyasi S. **Analysis of Occupational Accidents: A Data Mining Study.** Archives of Occupational Health. 2021; 5(3): 1051-8.

**Article History:** Received: 18 May 2021; Revised: 14 July 2021; Accepted: 23 August 2021

**Copyright:** ©2021 The Author(s); Published by Shahid Sadoughi University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

rate of accidents arising from work in developing countries.<sup>1,2</sup>

Various studies have identified different issues. They include poor organization of safety and labor, the size and greatness of industry, lack of coordination, pressure and time constraints, financial and budgetary constraints, lack of standardized data and information, poor internal and external communication, poor involvement of workers on safety issues, workers' expertise, poor training and failure, fatigue and wear, improper selection of equipment, wrong use or improper inspection of them, unconsciousness or poor safety awareness of senior management and project managers, The lack of preventive or protective equipment and some other factors have been identified as causes and factors affecting safety performance that can be used to conduct a comprehensive study.<sup>3-5</sup>

It should be noted that the difficulty of identifying, measuring and analyzing causes of events can be due to the difficulty in recording events data as well as the understanding of causal relation between the factors of incident makers and accident. However, the relationship between these causes as an important issue has to be determined because overall budget industries have limited safety and better results are achieved when an industry can identify and adopt the best safety strategy for the allocation of available resources.<sup>6-7</sup> Therefore, in regard to retrospective methods of accident analysis, it is possible to identify some causes of incidents and inefficiency of these methods and techniques in prediction and prevention of accidents, as well as the complexity of industrial environments and happening lots of events that it can cause losses and adverse effects even at the national and international level, preventing re-happening accidents and also new accidents will require to present a way to analyze and predict events. There is a need to prevent accidents due to their extreme consequences that require knowledge of the causal factors of the

accident, how the factors contribute to occurrence of the accident, the extent to which these factors contribute to happening the event, and the risk caused by these factors. The demands to get this knowledge led to the studies of the causality of events that Heinrich founded it in 1936. Beyond the contributions that have been achieved by various studies in the field of causality for the concept and understanding of events, these studies also depict the complex nature and the reasons why any accident occurs. Some studies have also shown that damage caused by accidents and occupational diseases affect not only safety and health but also the economy because of high costs associated with different types of damages and occupational injuries. On this basis, identifying the influential, participatory and related factors of work-caused injuries is an important and critical issue in the analysis of accidents and a way to prevent severe and catastrophic events.<sup>8</sup>

Studies have done a little on the influence of parameters associated with the work and determining the portion and extent of the effect of each of these parameters, as well as the interaction effects of these parameters on the country level.<sup>9</sup> No study has ever been done on certain geographical area where industrial and productive sectors operate under an administrative and governing conditions. Identifying causes and factors affecting the creation of events is an important and fundamental issue in preventing accidents. One of the most important tools related to the prevention of industrial accidents is descriptive - analytical analysis in the field of events that have been explained to understand the effective, influential and contributing factors in the incidence of occupational accidents. Extensive efforts have been made by researchers in different fields to explain different types of accidents and factors affecting them. These studies can explain and analyze while revealing the causes of accidents, job events, and eventually result in an understanding and prediction of events.<sup>10,11</sup>

According to what has been stated, this study aims to comprehensively analyze work-related accidents, identify parameters and variables related to these accidents and determine the effectiveness of each of these variables. It is designed in order to provide model and pattern for analysis of accidents caused by traumatic work and improving the safety and health situation in industries and studied sections. This study will analyze and determine the reasons for the causes of incidents resulting from work and related factors affecting the accidents during a 3 - year study.

## Methods

This study is a retrospective study of the descriptive - analytical kind aimed at analyzing and modeling industrial accidents in a three - year period (2017-2019) to provide a model for determining factors and factors affecting the accidents caused by work in Tehran province. The study population consisted of all incidents that occurred during the three - year period in part of the industries and working environments of Tehran province (4 districts of Tehran).

According to the community with debilitating and traumatic accidents, all accidents caused by traumatic work were considered and used in the sample case study. The initial sample volume in this method included 842 job accidents that had to be investigated and evaluated for all study criteria. The main variable was the study of debilitating and traumatic accidents that occurred from 2017 to 2019 in the studied industries. It is to be noted the only events remaining in the study were those in which all the investigated factors and variables were reported or the research team managed to obtain data from the collection process. The events were selected for the final study of mortality and all the debilitating incidents that have all the investigated variables. For example, incidents involving incomplete information in terms of demographic personal factors, or lack of an accident reporting sheet containing information about unsafe situations and activities as well as the type of accident were

excluded. It should be mentioned that, in the end, according to 842 cases of the collected work, 818 incidents of were eligible for study as a final sample.

### The implementation stage of the study

The implementation stages of this study included "five steps".

In the first stage, data related to the accidents caused by the work of the last three years in industries were studied and collected by referring to the number of accident data banks that were recorded in the job administration.

In this stage, data collection was carried out using a checklist of incidents, different reports of incidents and records of all incidents resulting from work over the past three years.

In the second step , screening and classification of all sorts of accidents caused by the work of the last three years in industries were studied and the required information were from them. At this stage, initially based on entry and exit criteria of study, the most important screened and identified accidents and then selected accidents were classified based on criteria and factors such as type of accident, site occurrence, severity of the accident, etc. ... and then the information were extracted.

The third step of this study includes the study and classification of different dimensions of factors and factors affecting the accidents resulting from work case study including the set of causes. At this stage, different aspects of the events were investigated and then the data obtained from IBM SPSS version 0.22 was released.

The fourth step: in this step, based on the objectives of the study, the regression analysis approach of these accidents was data mined.

### Data Analysis

In the collected data analysis from IBM SPSS was used for the nominal multi - system logistic regression analysis using the forward wald method, and IBM SPSS Modeler version 14.2 was used to select the important and superior features affecting

the work - related events. The statistical tests used in this study were two - way and were considered significantly lower than 0.05, which is worth mentioning.

## Results

### descriptive findings

The descriptive results of the injured people showed that the mean age and experience of injured workers were  $55/11 \pm 55/34$  and  $87/9 \pm 14/12$ , respectively. The descriptive results of the personality characteristics of being married people or education level of injured people indicated that approximately 1/3 of the injured ones were single and 2/3 of them married (single 3.31% and married 7.68 %). Moreover, almost 2/3 of injured people had undergraduate education (62.7%) and less than 10% of the injured people had university and academic education (3.7% associate's degree and 4.1% bachelor's degree and above). The descriptive findings of the factors related to the job in this study included the type of main activity of the workshop, the number of workers in each workshops and the type of job in injured people. Most accidents occurred which belonged to ,respectively, construction workshops (52.4 %), industry - production (24.4%) and other public and social services activities (11.1%).

The results of the frequency study of the number of workers in the event - related workshops based on the non-parameter Kolmogorov - Smirnov test shows that the distribution of frequency in the population groups was approximately uniform ( $p=0.07$ ). Moreover, the findings indicated that the highest percentage of accidents occurred for the small workshops of one to ten (40.5 %) and the lowest percentage of accidents occurred for the large workshops having more than 200 workers (5.5%) (Fig. 1).

Results related to the study of the type of job in incidents indicated that the maximum percentage of accidents occurred in the sequence of the operators/construction workers (45.3 %), the

operator / production worker (22.9 %) and different service professions (13.1 %) respectively. The lowest percentage of incidents occurred for cooking jobs and restaurant (5.1 %), drivers (5.3%) and installation and maintenance jobs (8.3%) (Table 1).

The results of the incident time parameter survey in day showed that the highest percentage of accidents occurred in shifts as in the morning (6-14), evening (14-20), and night (6-20) shifts, including 49.7 %, 40.6 % and 9.7 % respectively.

The results of the parameter investigation of the accident factors in the case study indicated that the accident factors in the events in the case study included 8 factors of the fall of objects, trapped inside or between objects or cars, contact with objects and hot surfaces, falling and slipping, collapsing and staying under debris, dealing with objects and equipment, electricity flow and other factors (fire, suffocation, contact with chemicals, poisoning). The highest percentage of accidents occurred respectively as falling factors and trips (49.1 %), dealing with objects and equipment (19.4%), trapped inside or between objects or cars (13.3 %) and the fall of objects (10.8%) (Table 2).

The causes of the accident were classified based on regulations as the lack of supervision on the workshop by the employer, lack of training the workers, lack of personal protective equipment, non - protection or lack of technical defect device, an unsafe procedure and negligence or carelessness of worker, non - use of personal protective equipment and not observance of safety issues. The highest percentage of accidents in the event were due to the causes of non - compliance of safety items in regulations (84.1%), no monitoring of the workshop by the employer (15.9 %) and acting in an unsafe procedure or negligent or careless employee (15.4 %), respectively. According to the table's findings, the lowest percentage of accidents occurred, respectively, was due to the causes of non - use of personal protective equipment (1.1 %), non-training

of workers (3.5 %) and machine without protection or having a technical defect (Table 3).

The highest percentage of accidents occurred in the result of the incident, including fracture (44.4%), injury (16.1 %) and death (13.3%), respectively. According to these findings, the lowest percentage of accidents occurred in the result of the accident, including the disability (4.3 %), burn (4.6 %) and amputation (3.5 %).

**The analytical findings based on the "selection of attributes" algorithm.**

The results of the most important variables affecting the outcome of the accident parameters as a job accident index using analytic analysis of Pearson  $\chi^2$  showed that the mean age, work experience, the number of workers, the accident time on day, factors of the accident with importance degree equaled 0.1 and two factors of education and marriage with a importance degree respectively equaled 0.988 and 0.955 ,qualified for entering into analysis and modeling the effects of these variables on the result of occupational accidents.

**The analytical findings based on "nominal multiple logistic regression analysis".**

The findings of the analysis and regression modeling factors affecting job accidents in terms of the accident result in Table 4 showed that in the final model of the seven working history variables, education level, type of workshop activity, the number of workers, the accident time of the day, the accident factors and the causes of the accident remained in the final model and were significant with the result of the incident ( $p < 0.05$ ).

The good evaluation of fitting the model using cox and cnell test showed that the  $R^2$  coefficient of the model was estimated at 0.72. The good evaluation of fitting the model using cox and cnell test showed that the  $R^2$  coefficient of the model was estimated at 0.72. The B values (correlation coefficient) indicated that the highest correlation value with the outcome of the incident related to the

variables of education value ( $B=3.03$ ) and work experience ( $B=1.56$ ) and the minimum correlation with the outcome of the event related to the type of workshop activity variable ( $B=1.09$ ) and the number of workers ( $B=1.13$ ), ( $p < 0.05$ ).

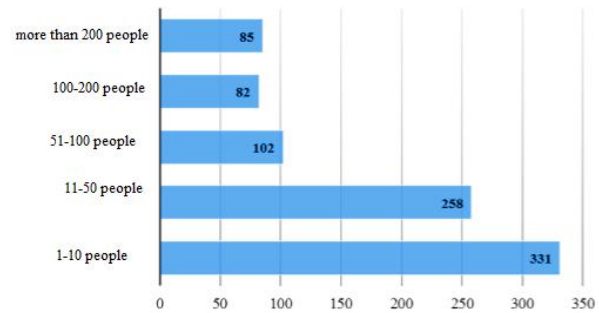


Figure 1: The results of the study of the frequency of workers in the workshops with accidents (n=818)

Table 1: Results of the study incidents based on job type (n=818)

Type of Job	Frequency (%)
Operator/Production Worker	187 (22.9%)
Operator/Building Worker	371 (45.3%)
Cooking & Restaurant	42 (5.1%)
Service	107 (13.1%)
Driver	43 (5.3%)
Installation and repairs	68 (8.3%)

Table 2. Findings of the parameter of incident factors in the studied incidents (n=818)

Causes of incident	Frequency (%)
Falling objects	88 (10.8%)
Getting stuck in or between objects or cars	109 (13.03%)
Contact with objects and hot surfaces	6 (0.07%)
Falling and slipping	402 (49.1%)
Spilling and staying under the oar	8 (1.0%)
Dealing with objects and equipment	159 (19.4%)
Electricity current	22 (2.7%)
Other	24 (2.8%)

Table 3. Findings of the causes of incident in the studied incidents (n=818)

Causes of event	Frequency (%)
Failure to supervise the workshop by the employer	130 (15.9%)
Non-training of the worker	29 (3.5%)
Failure to provide personal protective equipment	121 (14.8%)
Unprocessed or technically defective device	38 (4.6%)
Acting insecurely or negligently by the worker	126 (15.4%)
Not using personal protective equipment	9 (1.1%)
Non-compliance with safety requirements in accordance with regulations	688 (84.1%)

**Table 4.** Results of regression modeling of factors affecting incident based on accident outcome (n=818)

The remaining variable in the model	B	SE	† p-value	R <sup>2</sup>
Work Experience	1.56	0.07	0.001	0.72
Education Level	3.03	0.05	0.013	
Type of workshop activity	1.09	0.7	0.001	
Number of workers	1.13	0.3	0.001	
Time of incident per day	1.14	0.8	0.001	
Causes of incident	1.22	0.07	0.001	
Causes of incident	1.33	0.12	0.001	

## Discussion

This study aimed at data mining work-related accidents in a 3-year study based on retrospective data in four regions of Tehran. In this research, which utilized two data mining logics involving characteristic selection and nominal multiple logistic regression analysis, it was found that 818 continuous traumatic events during the five-year period from 2017 to 2019 were influenced by various factors, as well as the resulting parameter and consequences of these work-related accidents can be influenced by various parameters and factors and severity. They are caused by the role of these aggravated or reduced factors.

Pearson analysis  $\chi^2$  led to choosing the most significant variables influencing the accident outcome parameter as an indicator of occupational accidents based on logic and characteristic selection algorithm. These results indicated that 9 variables were chosen as the most important variables affecting this parameter of events. Furthermore, the results of regression analysis and modeling of factors influencing occupational accidents based on the result of the accident also indicated that in the final model, 9 variables of age, work experience, type of workshop activity, number of workers, time of accident per day, causes of accident and causes of accident, marital and education in the final model remained and had a significant relationship with the result of the accident ( $p < 0.05$ ).

The location and time of accident is a very significant factor to predict and analyze accidents.

Some research indicated that occupational accidents and injuries commonly occur during early shift periods due to lack of mastery of work and being in the shift phase and late shift work due to fatigue caused by the final hours of work.<sup>12, 13</sup>

A number of studies have indicated that inadequate, inefficient and inappropriate education can result in carelessness, dangerous behaviors and kinds of human error and influence the incidence and severity of accidents in the industry. Some research indicated that risk identification and risk perception can be enhanced by educational interventions.<sup>14-16</sup>

Factors related to job, kind of activity and work and factors related to the management structure of workplaces are one of the basic factors in the process of analysis and cause of accidents, which can have various roles in the incidence and severity of occupational accidents. These factors can cause problems in the process of safe work and implementation, or are influenced by interaction with other underlying factors indirectly in occupational accidents.<sup>17, 18</sup> In the analysis process based on two approaches of selecting characteristics and analytical multivariate regression in this research, the type of job and activity that led to accidents, the number of workers in the industry and the type of industrial workshop had a significant relationship with the accidents and occupational injuries.

One of the most significant factors involved in occupational accidents in the industry involves the characteristics of project and industry which can involve variables such as organizational, physical and operational characteristics of industry. Despite the importance of surveying accidents caused by these variables, the multi-cause characteristic of this phenomenon remains difficult and therefore more investigation is required.<sup>19</sup>

The results of this research indicated that the factor of the kind of accident and the conditions resulting from the accidents and the resulting

injuries, in addition to variables such as the type and manner of accidents such as falling from height, throwing and falling objects from height and hitting the worker, slipping and falling or getting stuck between objects and equipment, collision and collision, splurge of compounds and chemicals, contact with Electrical objects or circuits and accidents caused by the movement of loads and materials had a direct and significant relationship with accident indicators including the result of the accident and also the type of damaged organ caused by accidents in this research.

Thus, these factors and variables are regarded as a significant parameter in occupational accidents due to their direct role in the outcome of accidents and have an important role in identifying and analyzing the factors and causal factors of occupational accidents in the industry.<sup>20,21</sup> Although in this research, attempts were done to analyze and assess the accidents with a comprehensive approach, like other studies, this study has limitations such as lack of collecting all the required variables for a comprehensive analysis and examining all the variables affecting occupational accidents.

## Conclusion

The results provided from using feature selection algorithm and regression modeling indicated that the results of continuous accidents in the three-year period were influenced by multiple factors. These results indicated that in this research, data mining of 818 occupational and work-related accidents in different wards investigated the result of the accident included death and death of workers, amputation and amputation of the injured people, fractures in different parts of their bodies, burns and injuries caused by accidents and other outcomes except for the mentioned consequences under the influence of 9 important factors involving the mean age and workers' work experience, their education, the status of being married and single, and the type of activity which was studied in the workshop or industry, the

number of workers in each workshop or industry, the time of the accident per day, the causes of the accident and the causes of the accident .

## Conflict of interest

All authors declare they have no conflict of interest in relation to this study.

## Acknowledgement

This article is derived from a master's thesis (Code: 75520) in Health, Safety and Environment Management (HSE) at Islamic Azad University, Central Tehran University.

## Authors Contribution

Conceptual design: Ahmad Soltanzadeh; Samira Ghiyasi

Data gathering: Vahideh Dadfarma

Data analysis and modelling: Ahmad Soltanzadeh; Vahideh Dadfarma

Manuscript preparation: Ahmad Soltanzadeh; Vahideh Dadfarma; Samira Ghiyasi

## Reference

1. Ghanbari M, Ashtarian H, Yarmohammadi H. An investigation of the frequency of the occupational accident in Kermanshah, Iran (2009–2013). *Annals of Tropical Medicine and Public Health*. 2017;10(5):1306.
2. Soltanzadeh A, Mahdinia M. Path analysis of occupational injuries based on the structural equation modeling approach: a retrospective study in the construction industry. *Iran Occupational Health*. 2019;16(3):47-57.
3. Monazzam M, Soltanzadeh A. The relationship between the worker's safety attitude and the registered accidents. *Journal of research in health sciences*. 2009;9(1):17-20.
4. Soltanzadeh A, Mohammadfam I, Moghimbeygi A, Ghiasvand R. Exploring causal factors on the severity rate of occupational accidents in construction worksites. *International journal of civil engineering*. 2017;15(7):959-65.
5. Soltanzadeh A, Mohammadfam I, Moghimbeigi A, Ghiasvand R. Key factors contributing to accident severity rate in construction industry in Iran: a regression modelling approach/Primjena regresijskog modela u analizi ključnih čimbenika koji pridonose težini nesreća u građevinskoj industriji u Iranu. *Archives of Industrial Hygiene and Toxicology*. 2016;67(1):47-53.
6. Farshadnia SM, Arghami S, ShahabSafa A, Majidi F. A Budgeting Model for the Safety Unit of an Under Construction Metro Station



- in Tehran Using a Robust Optimization. *Journal of Human, Environment and Health Promotion*. 2018;4(3):131-7.
7. Alahmad B. High assurance real-time systems: scheduling, budgeting, and safety: University of British Columbia; 2019.
  8. Ashley S, Vaughan G, Nuttall W, Thomas P, Higgins N. Predicting the cost of the consequences of a large nuclear accident in the UK. *Process Safety and Environmental Protection*. 2017;112:96-113.
  9. Ghorbani A, Soltanzadeh A. An Investigation of the Attitudes of Health & Safety Personnel toward Safety in Construction Projects. *Archives of Occupational Health*. 2019;3(1):291-8.
  10. Alizadeh Savareh B, Mahdinia M, Ghiyasi S, Rahimi J, Soltanzadeh A. Accident Modeling in Small-scale Construction Projects Based on Artificial Neural Networks. *Journal of Human Environment and Health Promotion*. 2019;5(3):121-6.
  11. Mohammadfam I, Soltanzadeh A, Arsang-Jang S, Mohammadi H. Structural Equation Modeling Modeling (SEM) of Occupational Accidents Size Based on Risk Management Factors; A Field Study in Process Industries. *Health Scope*. 2018;8(1).
  12. Soltanzadeh A, Mohammadfam I, MOGHIM BA, Akbarzadeh M. Studying disabling occupational accidents in the construction industry during two years. 2014.
  13. Soltanzadeh A, Heidari H, Mohammad H, Mohammadbeigi A, Sarsangi V, Darakhshan Jazari M. Comprehensive causal analysis of occupational accidents' severity in the chemical industries; A field study based on feature selection and multiple linear regression techniques. *Health and Safety at Work*. 2019;9(4):290-9.
  14. Perlman A, Sacks R, Barak R. Hazard recognition and risk perception in construction. *Safety science*. 2014;64:22-31.
  15. Laal F, Moradi Hanifi S, Mohammadian Y, Fallah Madyari R. Investigating the Causes of the Incidents and Monitoring the Safety System before and after the Implementation of OHSAS 18001 in a Combined Cycle Power Plant. *Archives of Occupational Health*. 2018;2(4):233-9.
  16. Beheshti MH, Firoozi Chahak A, Hajizadeh R. Association between risk-taking behaviors and safety culture in construction workers. *Archives of Occupational Health*. 2017;1(1):13-7.
  17. Leveson N. A new accident model for engineering safer systems. *Safety science*. 2004;42(4):237-70.
  18. Mohammadi H, Heidari H, Arsang Jang S, Ghafourian M, Soltanzadeh A. Relationship between Reactive and proactive Safety Indices: A Case Study in the Chemical Industries. *Archives of Occupational Health*. 2020;4(4):842-8.
  19. Manu P, Ankrah N, Proverbs D, Suresh S. An approach for determining the extent of contribution of construction project features to accident causation. *Safety Science*. 2010;48(6): 687-92.
  20. Sarkar S, Ejaz N, Kumar M, Maiti J. Root Cause Analysis of Incidents Using Text Clustering and Classification Algorithms. *Proceedings of ICETIT 2019*: Springer; 2020. p. 707-18.
  21. Xu Q, Xu K. Analysis of the characteristics of fatal accidents in the construction industry in China based on statistical data. *International journal of environmental research and public health*. 2021;18(4):2162.