Extraction and analysis of construction safety hazard factors from open data

N K A H Rupasinghe\textsuperscript{1} and K Panuwatwanich\textsuperscript{1}

\textsuperscript{1}School of Civil Engineering and Technology, Sirindhorn International Institute of Technology, Thammasat University, Pathumthani, Thailand.

Email: kriengsak@siit.tu.ac.th

Abstract. Construction is one of the most injury prone industries worldwide. Concerns of health and safety of the employees in construction sites have been a vastly discussed topic for decades. In many countries, companies are required to report safety incidents by using catastrophe investigating report in their workplaces to relevant authorities whereby such data is made publicly available under the open data policy. These open datasets may be well structured or may require further preparation in order to be usable. Some datasets are in the form of reports, which require qualitative, textual analysis to extract insightful information. The purpose of this study is to extract safety hazard factors from an open dataset obtained from the US Occupational Health and Safety Administration, and to further analyse such factors using statistical analysis techniques. For each reported case, text analysis was carried out with the narrative data field describing the circumstances leading to safety incidents to extract safety hazard factors. These hazard factors were categorized into human factors, technical factors, external environmental factors, organizational factors and other factors. The results showed that hazards related to human factors are most common. Descriptive statistics also showed that the most frequent nature of accident was fractures and most frequently occurring accident event was falls to the lower levels. Such information can help to provide insights into the accidents occurred and how relevant authorities may devise strategies to improve construction site safety.

1. Introduction

Construction industry plays a major role in developing the economy of many countries [1]. It provides the required infrastructure facilities for other parts of the economy to develop by reflecting the economic development of the countries [2]. However, construction industry has been identified as one of the most injury prone industries in terms of severe injury, loss of work time [3-10], hospitalization [11], disability [12] and fatality [13]. According to the records of International Labour Organization [14], 2.3 million labours around the world, including both men and women were subjected to construction site accidents or diseases every year. Hence, concerns of health and safety of the employees in construction sites have become a vastly discussed topic amidst the researchers over decades.

In many countries, companies are required to report the data on health and safety related incidents in their workplaces to relevant authorities such as The Department of Labour and Occupational Health and Safety Organization through catastrophe investigating report which consists of a complete description of the accident event. Some of the countries have made these data publicly available under open data policy. These open datasets may be well structured or may require further preparation in order to be usable. Some datasets are in the form of reports, which require qualitative, textual analysis to extract...
insightful information. Recently, it seems that the text mining is becoming a popular tool among construction researchers to extract data. However, hazards identification in open data through text mining and natural language processing is quite challenging and this study tries to fill this gap by being a reference to such study.

The above intention suggested the purpose of this study, which is to extract safety hazard factors from an open dataset obtained from the US Occupational Health and safety Administration, and to further analyse such factors using statistical analysis techniques. Outcomes of this study provides the most common safety hazards factors, natures of accidents, accident events and activities which are prone to cause accidents.

2. Theoretical background

2.1. Construction safety statistics

As far as the statistics are concerned, out of 7 billion people 43% of people are employed [15]. Employment in industry accounts for 22.927% of total employment [16], but accounts for 58.2% of work place accidents [17]. The U.S. Construction industry employed 10.7 million people in United States [18] but records 21% of fatalities [19]. Such serious accidents lead to bad reputation of the construction industry as well as trigger further risks such as financial difficulties, project failures, time overruns etc. [20]. Since the construction injuries have a notable impact on the financial success of the construction industry, accidents can drastically increase the overall cost of construction up to 15% [21].

2.2. Safety hazards factors

Hazard can be identified as a specific situation connected to an activity and is characterized by such a configuration or state of factors of this process which may results in an accident at a work or occupational disease [21]. The health and safety commission (1995) defines hazard as “potential to cause harm”. Five main sources of hazards were identified by a previous research and they are presented in Table 1 below [22]. These five main hazard factors were used as a reference to subsequently extract hazard factors inherent in the taken open data set to achieve above mentioned objectives.

2.3. Common activities prone to cause construction accidents

It has been identified main hazardous activities which cause for the construction site accidents related to labourers by using total of 3,700 injured workers [23]. There were 16 hazardous activities and these activities were taken into consideration for classifying open data into meaningful information. List of those 16 activities were provided below.

| Sources of hazard          | Description                                                                 |
|----------------------------|-----------------------------------------------------------------------------|
| Human factors (TF)         | Human factors refer to ‘environmental, organizational and job factors, and human and individual characteristics which influence behavior at work in a way which can affect health and safety’ [24] |
| Organizational factors (OF)| Accidents occur due to safety culture, work planning, scheduling, supervision and/or procedures |
| Technical factors (HF)     | Equipment or material which contributes to accidents while operating inside the construction industry |
| External environmental factors (EEF) | Accidents occur due to environmental behaviours which are not under control of any other party |
| Other civil structures (OtherF) | Accidents occur due to other civil structures situated inside or outside the building site |
3. Research methodology
This study was carried out in three main stages. The first stage involved reviewing of literature in construction safety to identify the gaps in construction safety research areas. In the second stage, descriptive statistics related to the data set was obtained and in the third stage manual text analysis was carried out to identify sources of hazards at construction sites and hazardous activities. The final stage was to develop Pareto diagram to prioritise the most critical hazards and activities and refine the critical factors. The following sub sections provides the details of data collection and analysis.

3.1. Data collection
Data collection for this study was carried out with severe injury accident data set obtained through Occupational Health and safety Administration, Department of Labour, United States (Occupational Health and safety Administration) which was publicly available for downloading under open data policy. There were 33,018 number of accidents recorded from 1st January 2015 to 31st March 2018. Not only the construction site accidents, but also the accidents occurred in other activities, such as agriculture, mining, utilities, manufacturing, transportation, and educational services. The downloaded data was categorized according to event date, company or the employer, address, primary NAICS (North American Industry Classification System), severity, nature of accidents, part of the body damaged, cause of the accident, final narrative and secondary source of accident.

3.2. Descriptive analysis of data
Descriptive data analysis was done by using IBM SPSS statistical 23 software. Data analysis were carried out in several steps as shown in Figure 1. In the first step, construction related data were separated from the data set with the help of a macro function in Microsoft Excel using the primary NAICS. NAICS is the North American industry classification system which classifies the registered industries according to the type of business. According to the NAICS, codes begin with “23” were categorized as construction. In the second step, data were sorted according to descending order and graphs were plotted to identify common nature of accidents, body parts damaged and the events which causes the accidents with the help of another Excel macro. In the third step, most common ten events were identified and then nature of the accident occurs and parts of the body damaged due to each particular event was identified by using an Excel macro. Finally, SPSS was used to plot pie charts and depicts the results.

3.3. Text analysis of data
Final narrative which describes the accident event was used in this stage to identify sources of hazards and hazardous activities. As an example, “A worker was installing a retaining wall next to Highway 285 when a gust of wind displaced a tree next to the highway. The tree struck the worker, resulting in back, pelvis, and ankle injuries.” In this incident, hazardous activity is a “natural event” as the gust of wind displaced the tree and source of hazard is “external environment” which is not under control of any human. Likewise, 200 data from this final narrative data column was randomly taken and manual
text analytics were carried in order to identify sources of hazards and hazardous activities. Pareto charts was then drawn to prioritise the most critical factors and refine them.

4. Results and discussions

4.1. Descriptive statistics

4.1.1. Construction site accident events. Out of 214 different types of accident events, Figure 2 shows the graph plotted using all the accidents events and enlarged view of the first ten commonly occurred accident events. “Falls to different lower levels” is the main cause for many accidents occurred in construction sites, account for 39%. This is evident in the Figure 2 as the five places out of ten common accident events were taken by falls.

![Figure 1. Methodology of descriptive statistics analysis.](image)

![Figure 2. Popular events Vs frequency.](image)
4.1.2. Nature of the injury or the disability due to construction accidents. According to Figure 3, out of 103 different injuries, fractures are the most common type of injury and 38% of the accidents cause different kinds of fractures. Amputations are the second highest nature of accidents and 16% of total data set was accounted for by amputations.

![Figure 3. Nature of accident Vs frequency.](image)

4.1.3. Nature of accidents occur due to each event. Figure 4 to Figure 13 describe the nature of each accident event. In Figure 4, 6, 8, 10 and 13 it clearly shows that “fractures” is the most commonly occurred accident nature when a fall to lower level accident occurs. According to Figure 4, fractures accounts for 53.1%, traumatic injuries and disorders unspecified takes 11.2%, soreness, pain, hurt and unspecified injury takes 11% and others takes below 5%. Figure 5 shows when an accident occurred due to compressed or pinched by shifting object or equipment, amputations occurs 68.6%. In Figure 6, Figure 8 and Figure 9, Figure 10 and Figure 13 fractures takes 58.95%, 56.65%, 61.15%, 46.51 and 49.61%, respectively.

![Figure 4. Nature of accidents occur due to other fall to lower level, unspecified.](image)

![Figure 5. Nature of accidents occur due to compressed or pinched by shifting objects or equipment.](image)
Figure 6. Nature of accidents occur due to other fall to lower level 6 to 10 feet.

Figure 7. Nature of accidents occur due to other fall to lower level less than 6 feet.

Figure 8. Nature of accident occur due to other fall to lower level 11 to 15 feet.

Figure 9. Nature of accidents occur due to injured by slipping or swinging object held by injured worker.

Figure 10. Nature of accidents occur due to struck by falling object or equipment, n.e.c.

Figure 11. Nature of accidents occur due to exposure to environmental heat.
4.2. Text analysis results

4.2.1. Pareto analysis for types of hazards. Pareto analysis is statistical technique which is useful to identify possible causes of actions among competing list of actions. In this paper, Pareto analysis was used to identify number of hazards factors that has highest contribution in causing construction site accidents. This helps project managers and relevant authorities to identify significant hazards and take necessary actions. The key technique used here is 80/20 rule, which suggests that focus should be given on those critical hazards that cause 80% of the accidents. According to the Pareto chart presented in Figure 14 and Table 2, human factors and technical factors contribute around 86% of construction site accidents among five identified factors. Organizational factors which was taken as accidents occur due to safety culture, work planning, scheduling and supervision have a 0% contribution in the analysed 200 incidents. Therefore, this suggests that, mitigating regular mistakes made by the workers in the construction sites and giving attention to mobile equipment and tools can reduce construction site accidents drastically.

4.2.2. Hazardous activities at construction sites. Following Figure 15 shows hazardous activities at construction sites and their occurrences as a percentage. Climbing, walking, running causes for 10% of accident out of 200 data and as this was taken from open data, 15% of obtained data set does not mentioned the hazardous activity.

5. Concluding remarks

The purpose of this paper was to extract safety hazard factors from an open dataset obtained from the US Occupational Health and safety Administration, and to further analyse such factors using statistical analysis techniques. Descriptive statistics shows that falls are the most common construction site accident and fractures are the highest nature of accident occur due to falls. Based on the findings from Pareto analysis, human factors and technical factors have 86% contribution for causing construction site accidents. Furthermore, results of manual text analysis to identify hazardous construction site activities
shows that climbing, walking and running causes for highest number of accidents. However, this can be used as a reference to analyse the data extracted through text mining and development of a technique to extract those sources of hazards and hazardous activities through text mining and natural language processing is the future work of this study.

![Pareto diagram for sources of hazards](image1)

**Figure 14.** Pareto diagram for sources of hazards.

![Hazardous activities at construction sites as a percentage](image2)

**Figure 15.** Hazardous activities at construction sites as a percentage.

| Sources of hazards                                      | Frequency | Cumulative % |
|---------------------------------------------------------|-----------|--------------|
| Human factor (HF)                                        | 123       | 67.21        |
| Technical factor (TF)                                    | 35        | 86.34        |
| External environment factors (EEF)                       | 16        | 95.1         |
| Other structures situated inside the building site (OtherF) | 9         | 100          |
| Organizational factors (OF)                              | 0         | 100          |
| Total                                                   | 183       |              |

**Table 2.** Cumulative frequencies of each hazard factor.

Cause not mentioned
Climbing, walking, running
Inadvertently went off
Malfunctioning of vehicle or equipment
Cutting
Operating mobile equipment
Manually lifting, carrying, moving
Mechanically lifting, carrying, moving items
Not clear
Assembling, disassembling, or moving scaffolding
Building or installing forms or parts of structure, walls, roof, floor, siding
Descending
General maintenance or clean-up
References

[1] Al Haadir S and Panuwatwanich k 2011 Critical success factors for safety program implementation among construction companies in Saudi Arabia Procedia Engineering, 14 148-155

[2] Coble R J and Haupt T C 1999 Construction safety in developing countries: implementation of health and safety on construction sites. Proc. of the 2nd Int. Conf. of int. Council for Research and Innovation in Building and Construction CIB Working Commission W99. Honolulu

[3] Brenner H and Ahern W 2000 Sickness absence and early retirement on health grounds in the construction industry in Ireland J. of Occupational and Environmental Medicine 57(9) 615-620

[4] Derr J Forst L Chen H Y and Conroy L 2001 Fatal falls in the US construction industry, 1990 to 1999. J. of Occupational and Environmental Medicine 43(10) 853-860

[5] Dong W E I Vaughan P Sullivan K and Fletcher T 1995 Mortality study of construction workers in the UK Int. J. of Epidemiology 24(4) 750-757

[6] Kartam N A Flood I and Koushki P 2000 Construction safety in Kuwait: issues, procedures, problems, and recommendations Safety Science 36(3) 163-184

[7] Kines P 2001 Occupational injury risk assessment using injury severity odds ratios: male falls from heights in the Danish construction industry, 1993-1999. Human and Ecological Risk Assessment 7(7) 1929-43

[8] Larsson T J and Field B 2002 The distribution of occupational injury risks in the Victorian construction industry Safety science 40(5) 439-456

[9] Loosemore M and Andonakis N 2007 Barriers to implementing OHS reforms–The experiences of small subcontractors in the Australian Construction Industry Int. J. of Project Management 25(6) 579-588

[10] Wadick P 2007 Safety culture among subcontractors in the NSW domestic housing industry J. of Occupational Health and Safety Australia and New Zealand 23(2) 143

[11] Spangenberg S Hannerz H and Tüchsen F 2005 Hospitalized injuries among bridge and tunnel construction workers Construction Management and Economics 23(3) 237-240

[12] Hannerz S Spangenberg S Tüchsen F and Albertsen K 2005 Disability retirement among former employees at the construction of the Great Belt Link Public Health 119(4) 301-304

[13] Tüchsen F Hannerz H and Spangenberg S 2005 Mortality and morbidity among bridge and tunnel construction workers who worked long hours and long days constructing the Great Belt Fixed Link Scandinavian J. of Work Environment and Health 31(2) 22-26

[14] World Statistics ILO.org 2009 [online] Available: https://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang--en/index.htm. [Accessed: 26/06/2019]

[15] L. Evans Global employment: What is the world employment rate? The Guardian 2019 [Online] Available: https://www.theguardian.com/news/datablog/2011/jan/25/global-economy
globalrecession. [Accessed: 26- Jun- 2019]

[16] “Employment in industry (% of total employment) (modelled ILO estimate) | Data” Data.worldbank.org 2019 [online] Available: https://data.worldbank.org/indicator/SL.IND.EMPL.ZS?end=2018&start=1991&view=chart. [Accessed: 26- Jun- 2019]

[17] Common Workplace Accident Statistics Aftermath.com 2019 [Online] Available: https://www.aftermath.com/content/workplace-accident-statistics/ [Accessed: 26- Jun- 2019]

[18] Wang T Construction sector employees United States 2017 | Statistic Statista 2019 [Online] Available: https://www.statista.com/statistics/187412/number-of-employees-in-us-construction/. [Accessed: 26- Jun- 2019]

[19] Construction Safety - 21% of All Work Deaths in the US were in Construction – IMEC Technologies IMEC Technologies 2019 [Online] Available: http://www.imectechnologies.com/2018/07/30/construction-safety-21-of-all-work-deaths-in-the-us-were-in-construction/ [Accessed: 26- Jun- 2019]
[20] Zou Y, Kiviniemi A and Jones S W 2017 Retrieving similar cases for construction project risk management using Natural Language Processing techniques *Automation in Construction* **80** 66-76

[21] Hoła B 2008 Qualitative and Quantitative Modelling of Accident Situation in Building Industry Wroclaw: Oficyna Wydawnicza Politechniki Wroclawskiej

[22] Hoła B 2010 Methodology of hazards identification in construction work course *J. of Civil Engineering and Management* **16(4)** 577-585

[23] Jannadi O A and Almishari S 2003 Risk assessment in construction *J of Construction Engineering and Management* **129(5)** 492-500

[24] Health and Safety Executive (HSE) 2009 Reducing error and influencing behavior Health and Safety Executive Books