Occupational accident analysis according to professionals of different construction phases using association rules

Abstract

Despite continuous efforts to improve safety, worker safety awareness on construction sites is a major concern as it remains one of the most dangerous industries. The large number of factors involved in accidents and the complexity of the relationships between them make management difficult for managers. Therefore, potential hazards cannot be identified in order to develop effective safety procedures. This study addresses this problem by using the association rule method of data mining to extract knowledge from historical data of construction accidents. It can help managers to identify and provide frequent conditions that can be prevented in future by controlling risks on site. Occupational accidents that have been notified through an official electronic system on Spanish construction sites between 2003 and 2015 are analysed. Data have been divided according to professionals to explore the accidents in each construction phase. The results show patterns and recurrent factors with multiple relationships in all phases of the construction works. This is the case for the outsourcing variable which is a critical factor in occupational accidents in all construction phases. Similarly, the results have also shown that risk assessment is not an insurmountable barrier to accidents. Consideration of the different stages of the work provides flexibility in order to improve risk reduction and control actions. The results of the study provide a framework for improving safety practices, providing a valuable reference for all agents involved in the construction industry to improve risk management, preventive measures and action plans.

Keywords: Construction, Occupational Accidents, Construction Phases, Association Rules, Professions

1. Introduction

The construction industry presents high rates of occupational accidents compared to other industries due to its dynamic, simultaneous, temporary, complex, and decentralised nature (Gao et al., 2016; Dong et al., 2018;

Mohammadi et al., 2018; Martínez-Aires et al., 2018; Poh et al., 2018; Loosemore & Malouf, 2019; Lozano-Díez et al., 2019). In Europe (EU-28), "more than one fifth of all fatal accidents at work in the EU-28 took place within the construction sector" according to the European Statistical Office (2019). Concretely, in Spain, in the last ten years, the incidence rate in the construction sector has been higher than other sectors. Note that the incidence rate is the relation between the number of accidents and the average number of people exposed to risk (Labour, Migrations and Social Security Ministry, 2019).

Despite continuous efforts to improve safety, the high rate of accidents in the construction industry remains a major concern for both practitioners and researchers (Umer et al., 2018; Newaz et al., 2019; Martínez-Rojas et al., 2020). In this regard, there is a need for further knowledge on causal factors in construction accidents in order to develop preventive strategies (Martínez Rojas et al., 2013). One possible way to increase knowledge of how accidents are caused is through the accident investigation technique. Research on existing data on accident causation has shown that the context in which a construction site is operated and its particular characteristics are relevant (Gunduz et al., 2016).

In Europe, Framework Directive 89/391/EEC on the introduction of measures to encourage improvements in the safety and health of workers at work introduced the obligation for employers to keep a list of occupational accidents that result in a worker being unfit for work for more than three working days. In Spain, these notifications must be sent through the electronic system "DELT@", which involves the completion of an official workplace incident notification form with a total of 58 variables according to the third edition of the methodology of the ESWA (Eurostat, 2013). These variables include factors of a more permanent nature, and also the scene and mechanism of the accident (deviation and contact).

However, knowledge of isolated accidents is not enough to draw conclusions about the characteristics of the accident rate phenomenon. Such conclusions can be drawn on the basis of results obtained from research on a considerably numerous sets of accidents (Hola & Szóstak, 2017a). Additionally, given that no single factor provides a complete explanation of the high incidence rate in the construction industry, it would be interesting to study the combined effect of multiple factors (Cheng et al., 2012).

For this purpose, it might be interesting to explore historical accidents which are available in the Spanish official database, concretely from 2003 to 2015. A comprehensive examination of the accidents included in the historical accident records would contribute to reducing the number and severity of accidents that occur in the construction workplace (Shin et al., 2018). In addition, it may be interesting to analyse the results from the perspective of the different types of work that are necessary during the execution process. To the best of our knowledge, there are no works in the literature that analyse the influential factors affecting occupational accidents according to the different occupations in the diverse phases of construction. For this purpose, in this article, four groups are proposed in order to analyse them separately in a first instance, and then analyse them together. The first group comprises the workers in the initial phase of the execution process: foundations and structure workers. The second group includes masonry and installations workers, while the third group considers workers engaged on the completion of the building. Finally, a fourth group includes workers involved throughout the whole execution process, such as managers and construction workers.

To analyse the mentioned data, there are well-established techniques such as data mining that are widely applied to analyse large amounts of data (Martínez-Rojas et al., 2021). Among the diverse techniques, the association rule mining technique is proposed, which allows the interpretation of interesting relationships among multiple factors (Guo et al., 2019). The association rule technique makes it possible to detect when the occurrence of a variable is associated with the occurrence of other variables in the same transaction, which in our case would correspond to an accident. The result of this analysis might help managers to prevent and to avoid the repetition of similar accidents in the future and to carry out risk control plans.

Therefore, the aim of this work is to explore occupational accidents in Spain between 2003 and 2015 from the official annual digital database. More specifically, this study seeks to identify crucial factors and occupational accident occurrence patterns in the construction industry, showing how these factors and patterns are in each phase of the construction work. To achieve this objective, the association rule mining technique is proposed, which allows finding hidden patterns and helping decision makers by showing the probability of relationships between data items within large data sets, such as the case of occupational accidents in the construction domain. The research developed in this work provides a better understanding of frequent and widespread unsafe acts among different types of workers during different construction phases that allows future prevention in certain occupations.

After this introduction, the remainder of the paper is structured as follows. Section 2 introduces some previous work from two perspectives: factors contributing to occupational accidents and proposals applying data mining techniques in this research domain. Section 3 explains the methodology in different steps. Firstly, the data collection and selection of data is explained. Then, the selection of variables is detailed and finally Section 4 introduces the association rule mining technique. Sections 5 and 6 present and discuss the results obtained while Section 7 presents the conclusions and guidelines for future research.

2. Background

Due to the impact that occupational accidents have in the construction sector, many authors have focused on the analysis of the main factors that influence this incidence. This section presents previous works from two different perspectives: firstly, from the perspective of recognised factors that have been identified by other authors as closely related to the causes of the accident; secondly, from the perspective of proposals that have applied data mining techniques in the context of this research work.

2.1. Factors contributing to occupational accidents

Accidents are defined as unforeseen events that result in injury, death, loss of production, and property damage. Over time, many researchers have tried to understand occupational accidents and illness in the construction industry by proposing accidents causation models and analyzing the elements contributing to occupational accidents and illnesses (Zhang et al., 2020; Sousa et al., 2014). The various existing accident causation models show some fundamental differences as can be observed in Figure 1.

	Sequential models	**Epidemiological models**	**Systemic models**
Basic principles	Causality (single or multiple causes)	Latent conditions. Hidden dependencies	Dynamic couplings. Functional resonance
Critique	Simplistic and omits data/interactions. Not value organizational or management factors	Based on inadequacies of the organization and management. Fails to directly evaluate causation	Accidents considered as a control problema. Difficult to represent graphically
Techniques	Fault Tree, Ishikawa diagram, Domino theory, SCAT	Barriers analysis, change analysis, ECFCA, RCA	STEP, TRIPOD BETA, FRAM

Fig. 1. Accident causation models, principles, critique, and techniques.

At first, workplace accidents have been understood according to sequential causal models or simple linear models such as Henrinch's known as the “domino theory” (Heinrich, 1941). In this model, an accident is the culmination of a series of events and circumstances, and the contribution of this model is to recognize that an accident can be avoided by removing any factor in the sequence leading to it. The following models consider accidents as a result of the convergence of multiple sequences of events. Within the epidemiological models, Reason's "Swiss cheese" model stands out (1997). This model guides the analysis by revealing the complex interaction of latent conditions, active faults, and barriers. Thus, although many unsafe acts occur, only a few materialize in accidents due to the existing barriers represented by the layers of a Swiss cheese. More recently, given the limitations of the previous models, other types of models have emerged: systemic models or non-linear models. Accidents are caused by unexpected combinations of normal actions, rather than human error (Pardo-Ferreira et al., 2018). Information from previous accidents help to plan responses to future disturbances to be more efficient. These models attempt to capture the dynamic interaction of environmental, cultural, organisational and other factors in creating a hazardous situation (Golizadeh et al., 2018).

In the literature, many authors have investigated the variables that contribute to the high rate of occupational construction accidents. In this section, these variables are grouped in sets that bring together aspects of a similar nature. Firstly, the personnel set includes factors regarding the worker himself/herself, such as length of service, age, occupation, etc. On the other hand, the company set refers to factors such as the number of workers, location, etc. The accident set includes factors directly related to the accident, such as the day and hour of the accident, the cause of the accident and so on. Finally, the project set refers to information about the company in which the injured worker is employed, including, for example, the budget or schedule.

As can be observed in Table 1, the variables of the personnel and accident sets are those that have been considered in the largest number of proposals. Also, variables concerning the company have been analysed, although only two factors have been identified: the company size and the activity of the company. The set that has received least attention is the one regarding the project information. This might be due to the difficulty encountered in accessing project information. Note that on the construction site the injured worker does not have to work for the company in charge of executing the project.

	Factors															
	Personnel						Company		Accident						Project	
	Age	Gender	Nationality	Occupation	Length of service	Contract	Company size	Code of Activities	Day of week	Hour of day	Type of activity	Deviation	Place of accident	Cause	Financial aspects	Project
Chau et al. 2004	x			x							x					
López et al. 2008	x			x			x		x	x	x					
Liao and Perng 2008																
Chi et al. 2009		x			x		x				x	x				
Fernández-Muñiz et al. 2009							x									
Cheng et al. 2010		x			x		x				x	x	x			
Rivas et al. 2011	x		x	x	x	x			x	x						
Arquillos et al. 2012	x				x			x					x			
Cheng et al. 2012	x	x		x		x	x				x			x		x
Silva and Jacinto 2012	x			x					x			x	x	x		
Nielsen 2014																
López-Arquillos et al. 2015	x			x	x		x						x			
Stoilkovska et al. 2015	x															
Kang et al. 2017											x					x
Rameezdeen and Elmualim 2017	x	x		x	x		x	x								
Hoła and Szóstak 2017	x	x			x	x	x				x					
Al-Bayati et al. 2018			x													
Wang et al. 2018																
Hasebe and Sakai 2018	x			x			x									
Winge and Albrechtsen 2018	x	x									x	x				x
Shao et al. 2019									x	x	x					
Zhang et al. 2019														x		
Marzouk and Enaba 2019						x										
Akboğa Kale 2020	x				x	x			x	x	x	x			x	x
Martínez-Rojas et al. 2021	x		x	x	x		x	x	x	x	x	x				
Shepherd et al. 2021			x													

Table 1
Factors analysed in the literature grouped in sets of variables.

The variables included in the personnel set that have been studied in the most proposals are age, occupation, and length of service. Note the low impact of the worker's nationality when analysing occupational accidents. Regarding the company set, the size of the company is the most studied, but some proposals take into consideration the activity of the company. Finally, concerning the accident set, most proposals focus on the type of activity the worker was doing at the moment of the accident. Other factors receive less attention.

This analysis of proposals will be a very useful support when deciding about the relevant variables to consider in our proposals, as long as they are contained in the official accident notifications.

Regarding the accident model, variables are classified as antecedent and consequent. Antecedent variables are not included in the accident mechanism, while consequent variables are included in the accident mechanisms, and comprise Physical Activity, Deviation and Contact (Carrillo-Castrillo et al., 2015).

The next section reviews the proposals in the literature from the perspective of the data mining technique that has been applied.

2.2. Data mining

As mentioned in the Introduction, data mining techniques make it possible to explore data from a large dataset. Additionally, they can transform data into a more easily understandable structure (Witten et al., 2016). Taking the desired objective into account, several approaches such as discovering unusual data (e.g., anomaly detection), finding relations among variables (e.g., association rule mining) and exploring groups of data (e.g., cluster analysis) are found.

In this work, the proposed objective is to explore large amounts of occupational accidents in order to find patterns hidden in the mass of data. To do this, the most appropriate technique is to find association rules, which have been successfully used in a variety of research domains, such as market basket analysis (Valle et al., 2018), the mining sector (Sanmiquel et al., 2015) or traffic (Xu et al., 2018). The construction industry has also been aware of the opportunities offered by this technique, and it has been applied to different problems, such as construction defects (Cheng et al., 2015), building design (Eastman et al., 2009), building performance (Xiao & Fan, 2014; Martínez-Rojas et al., 2018), etc. Nevertheless, in this section, we only focus on proposals that apply data mining to occupational safety analysis.

In the literature, as previously mentioned, the technique most applied is the association rule:

- Liao & Perng (2008) explored the patterns of occupational injuries in the construction industry.

- Cheng et al. (2010b) performed an analysis of 1347 accidents in the Taiwan construction industry during the period 2000–2007.

- Li et al. (2017) applied association rules to find a relationship between the contributing factors and non-helmet use behaviours.

- Shin et al. (2018) explored intuitive knowledge expressed as association rules from a database of 98,189 serious injuries and fatal accidents that occurred in Korea during the period 2006–2010.

- Wang et al. (2018) proposed the use of association rules to ameliorate workplace hazard identification performance.

- Martínez-Rojas et al. (2021) applied association rules to analyze occupational accidents taking into consideration the nationality of the workers.

Most authors have highlighted that the use of association rules provides more intuitive results than classical statistical techniques. This is due to the large number of factors involved and the complexity of the relationships between them. Association rules help managers to identify potential hazards and then develop effective safety procedures.

3. Methodology

This section details the methodology that has been applied in our proposal according to the following stages:

- Stage I: data collection
- Stage II: the selection of variables
- Stage III: the selection of data regarding construction from the entire datasets
- Stage IV: the division of data according to the occupations of each phase of the construction process.
- Stage V: the application of association rules
- Stage VI: the analysis of the obtained results

The general scheme of the proposed methodology is illustrated in Figs. 2 and 3.

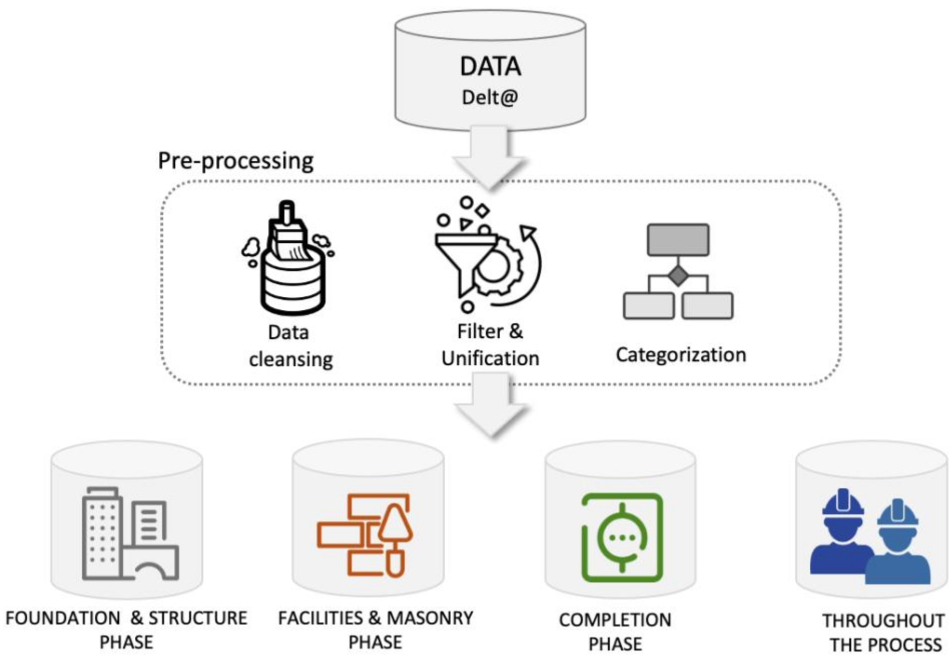


Fig. 2. The pre-processing process.

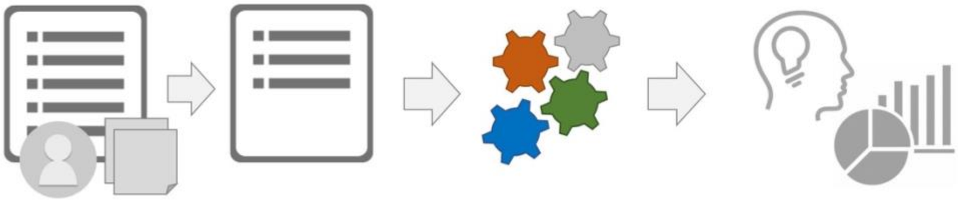


Fig. 3. Selection of variables and analysis of results.

3.1. Data collection

Some countries have repositories of occupational accidents, as in the case of Taiwan (Cheng et al., 2010a), Korea (Jo et al., 2017), Norway (Winge et al., 2019), etc. In Spain there is also a requirement that, from 2003, all accidents resulting in one or more days off work must be notified through the electronic system "DELT@" (del Estado, BOE). This process involves the completion of an official workplace incident notification form, where each accident is identified by 58 variables according to the third edition of the methodology of the ESAW (Eurostat, 2013). These variables refer to aspects of a different nature: details of the injured person, general and specific information about the accident, company data, etc.

3.2. Selection of variables

As mentioned before, the Spanish workplace incident notification form has a total of 58 variables. As a first step of our proposal, a selection of relevant variables for our study is carried out. This selection process takes two main criteria into consideration, namely relevant published results on this topic and our previous experience and analysis.

For this purpose, firstly, in Section 2 the variables that have been identified by the authors as relevant and contributing to occupational accidents were presented in Table 1. These variables were grouped in four categories: personnel, company, accident, and project. In this proposal, we include variables of all categories except those regarding the project, given that this kind of information does not appear in the mechanisms established for reporting accidents. This issue does not diminish the interest of the analysis since, as discussed above, this variable is very little analyzed in the literature. The reason is the difficulty in obtaining private information from the companies executing the projects. Nevertheless, other interesting variables have been identified that could be of interest for this study.

Secondly, an analysis to gain an overall understanding of the variables in our datasets regarding occupational accidents in different construction phases was performed. Some variables that appear in more than 80% of the cases are discarded because they detract from the relevance of the results, making interpretation of the results more difficult. For example, if 97 percent of accidents occur in people of male gender, too many rules that include this variable will be obtained. This will complicate the analysis and more interesting rules will not stand out. Therefore, these variables are removed after analyzing them in on a one-to-one basis for obtaining more focused results. This is the case of type of employment contract and Gender variables.

Description	Categorisation of Variables
Age of the victim at the time of the accident	**A1** (below 20 years), **A2** (21-30 years), **A3** (31-40 years), **A4** (41-50 years), **A5** (51-60 years), **A6** (over 60 years)
Occupation: victim's occupation at the time of the accident	**O1** (Managers and team leaders), **O2** (Heads of workshop and finishing workers), **O3** (Painter managers), **O4** (Workers in reinforced concrete), **O5** (Bricklayers and related works), **O6** (Carpenters), **O7** (Other structural construction workers), **O8** (Plasterers), **O9** (Plumbers and pipe fitters), **O10** (Floor layers and tile setters), **O11** (Painters and related works), **O12** (Building structure cleaners), **O13** (Air conditioning and refrigeration mechanics), **O14** (Other installers), **O15** (Electricians), **O16** (Construction labourers)
Length of service in the company	**LS1** (Less than 1 month), **LS2** (1-2 months), **LS3** (2-6 months), **LS4** (6 months-1 year), **LS5** (1-3 years), **LS6** (3-5 years), **LS7** (5-10 years), **LS8** (10-20 years), **LS9** (more than 20 years)
Size of enterprise: number of employees working at the local unit of the workplace	**S1** (1-9 employees), **S2** (10-49 employees), **S3** (50-249 employees), **S4** (250-499 employees), **S5** (500 employees or more)
Main **contractor** or **subcontractor**	**1** (Main contractor), **2** (Subcontractor)
Day of the week: day of the week when the accident occurred	**DW1** (Monday), **DW2** (Tuesday), **DW3** (Wednesday), **DW4** (Thursday), **DW5** (Friday), **DW6** (Saturday), **DW7** (Sunday)
Habitual work	**H1** (Yes, is habitual work), **H2** (No, is not habitual work)
Risk evaluation: If there is a risk assessment of the work in which the accident occurred	**R1** (Yes), **R2** (No)
Physical activity: the activity being performed by the injured worker just before the accident	**PA0** (No information), **PA1** (Operating machine - Not specified), **PA2** (Working with hand-held tools - Not specified), **PA3** (Driving/being on board a means of transport or handling equipment - Not specified), **PA4** (Handling of objects - Not specified), **PA5** (Carrying by hand - Not specified), **PA6** (Movement - Not specified), **PA7** (Presence - Not specified), **PA8** (Other specific physical activities not listed in this classification)

Deviation: a description of the way in which the circumstances of the accident differed from normal practice	**D0** (No information), **D1** (Deviation due to electrical problems, explosion, fire - Not specified), **D2** (Deviation by overflow, overturn, leak, flow, vaporisation, emission - Not specified), **D3** (Breakage, bursting, splitting, slipping, fall, collapse of material agent - Not specified), **D4** (Loss of control (total or partial) of machine, means of transport or handling equipment, hand-held tool, object, animal - Not specified), **D5** (Slipping - stumbling and falling - fall of persons - Not specified), **D6** (Body movement without any physical stress (generally leading to an external injury) - Not specified), **D7** (Body movement under or with physical stress (generally leading to an internal injury) - Not specified), **D8** (Shock, fright, violence, aggression, threat, presence - Not specified), **D9** (Other deviations not listed above in this classification.)
Contact: the contact that injured the victim	**FC1** (Contact with electrical voltage, temperature, hazardous substances), **FC2** (Drowned, buried, enveloped), **FC3** (Horizontal or vertical impact with or against a stationary object (the victim was in motion)), **FC4** (Struck by object in motion, collision with), **FC5** (Contact with sharp, pointed, rough, coarse material agent), **FC6** (Trapped, crushed, etc.), **FC7** (Physical or mental stress), **FC8** (Bites, kicks, etc.), **FC9** (Other contacts - modes of injury not listed)

Table 2 Categories of worker and accident variables

Once the variables have been selected, it is necessary to categorise them into groups. For this purpose, we have used the European Statistics on Accidents at Work system (Eurostat, 2013). Table 2 details both the variables defining the characteristics of the worker and the company involved in the accident and the variables related to the accident itself. Concerning the first group, the following variables are detailed: the age of the injured worker at the time of the accident, occupation and experience in months of the injured worker and the number of employees of the company. In connection with the second group: day of the week of the accident, if the work done is habitual, if there is risk evaluation and specific variables of the accident, such as deviation, physical activity or contact.

These latter variables provide valuable information about the time before the accident. It should be noted that these variables refer to the worker's own activity, without taking into account the environment in which the accident occurs. In the following, they are explained:

- Physical activity refers to the activity being performed by the injured worker at the exact time of the accident.
- Deviation describes the last event deviating from normality and leading to the accident. This is the most immediate cause that triggered the accident.
- Contact refers to the precise way in which the departure from normal practice resulted in an accident. It describes how the victim was hurt (physical or mental trauma), and thus represents the accident itself.

Having detailed the data preprocessing and filtering step and the selection of variables, the next section focuses on explaining the data selection process.

3.3. Selection of data

As mentioned before, this study considers occupational accidents in Spain from 2003 to 2015 in all sectors. The initial population is constituted by 5,495,609 instances of occupational accidents recorded for this period. The first objective is to obtain occupational accidents related to the construction sector.

To address this filtering process, the KoNstanz Information MinEr (KNIME) software has been applied. It allows large amounts of data to be managed and diverse filters to be applied in an easy and efficient way. In order to obtain the desired data, different filters need to be applied and it is also important to take changes in codification into consideration. For example, the occupation variable codification changed in 2011 and the variable regarding the economic activity of the company changed in 2009. After considering these changes of codification, it will be necessary to unify these variables to ensure all accidents are integrated.

After this process, a total of 1,525,865 accidents are retrieved, which represents 27.77% of all accidents. The categorisation of occupations is shown in Table 2.

3.4. Division of occupations according to construction phases

The execution of a construction site is divided into different phases (De Solminihac, 2011). This division is used for different purposes, including accident investigation (Carrillo-Castrillo et al., 2017; Salguero-Caparrós et al., 2015; Trillo-Cabello et al., 2020). For the purpose of this study, construction being a labour-intensive activity, we will divide its execution into three large groups: foundations and structure, installations and masonry and completion (Mínguez et al., 2004), identifying the main occupations involved in the different phases. A fourth group will be considered for those occupations involved throughout the construction process.

Once the construction occupational accidents have been filtered, the dataset will be split according to the four established phases based on the construction process: (1) foundation & structure phase (construction workers and craftsmen in foundation and structural construction works), (2) facilities and masonry phase (masons and installers in construction works), (3) completion phase (painters, carpenters and tilers) and (4) during the entire process (managers and construction Labourers). Table 3 shows the distribution of accidents in each of the phases and their percentage according to the total number of accidents. As can be observed, the phase including workers from facilities and masonry presents the largest number of accidents, with a total of 40%. Close to this percentage is the group of workers involved throughout the whole process, with a percentage of 30%.

Phase	Nº of accidents	Percentage
Phase 1	291.273	19%
Phase 2	619.612	41%
Phase 3	157.262	10%
Phase 4	457.718	30%

Table 3 Distribution of accidents per phase.

Occupation	Nº of accidents	Percentage
Managers and team leaders (O1)	30.756	7%
Heads of workshop and finishing workers (O2)	3929	1%
Construction labourers (O3)	423.033	92%

Table 4 Distribution of accidents in phase 4.

As previously explained, Phase 4 includes workers of different types that take part in the construction process in a continuous way. For this reason, it is worth disaggregating the number of accidents and percentages within this category. As can be observed in Table 4, 92% of accidents in this group correspond to construction labourers (O3) and only 8% correspond to managers or workshop heads. Therefore, in accordance with the results in Table 3, the most likely workers to suffer accidents are facilities and masonry workers and construction labourers.

4. Association rule mining

Association rules are a popular machine learning method for discovering insightful and interesting relations between variables in large datasets (Agrawal et al., 1993). More specifically, they are used to find correlations between different elements of the same event, i.e., to extract subsets of elements or attributes that appear frequently in the event and their correlation between them (Xu & Luo, 2021).

Generally, an association rule is expressed in the form of (A ⇒ B), where A is the antecedent part and B is the consequent part. Then, A and B represent the "If" part and the "Then" part, respectively. For example, a simple association rule extracted from the construction accident database could be:

$$Age < 30 \wedge Occupation = electrician \Rightarrow Accident = fatal \qquad (1)$$

This example rule would mean that an electrician who is aged "less than 30" usually tends to suffer serious accidents.

One of the most commonly used algorithms for finding frequent itemsets in a dataset for boolean association rules is the Apriori algorithm (Agrawal et al., 1994). On the one hand, the algorithm uses prior knowledge of frequent itemset properties, and, on the other hand, a rule is generated according to predefined minimum measures. Three measures are widely applied in the literature to quantitatively evaluate the obtained association rules: Support (S), Confidence (C) and Lift (L).

• Support (S) is defined as

$$S(A \Rightarrow B) = P(A \cup B), \qquad (2)$$

where P means probability, representing the probability that both itemsets A and B occur simultaneously in a transaction. Support is symmetric. Therefore, the support of rule A ⇒ B is equivalent to the support of B ⇒ A.

• Confidence (C) is defined as

$$C(A \Rightarrow B) = P(B \mid A) = \frac{P(A \cup B)}{P(A)}, \qquad (3)$$

representing the conditional probability that B is in a transaction where A is. It is not symmetric. Therefore, the confidence of the rule A⇒ B may be different from the confidence of the rule B ⇒A.

Given that support (S) and confidence (C) are probabilities, their values are in the interval [0, 1]. The association between two events is higher as these values are closer to 1.

• Lift (L) is defined as

$$L(A \Rightarrow B) = \frac{Confidence}{P(B)} = \frac{P(B|A)}{P(B)} = \frac{P(A \cup B)}{P(A)P(B)}, \qquad (4)$$

which measures how many times more often A and B occur together in a transaction than would be expected if their occurrences were statistically independent.

This measure can be interpreted as detailed below:

- L = 1 indicates no correlation between antecedent and consequent.
- L > 1 indicates positive correlation between antecedent and consequent.

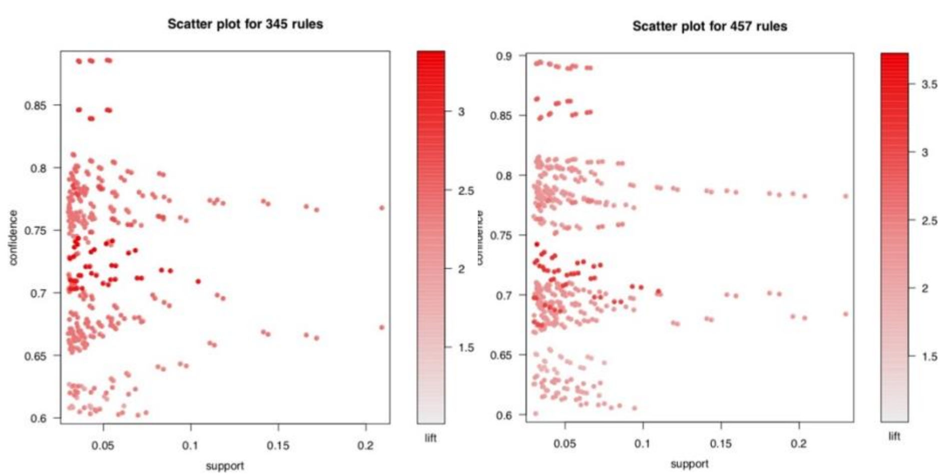
- L < 1 indicates negative correlation between antecedent and consequent.

Having explained the methodology, the next section presents the results obtained as well as a detailed discussion of these results. Note that the Apriori algorithm has been applied for each dataset separately, with the aim of comparing the results among the different occupations.

5. Results

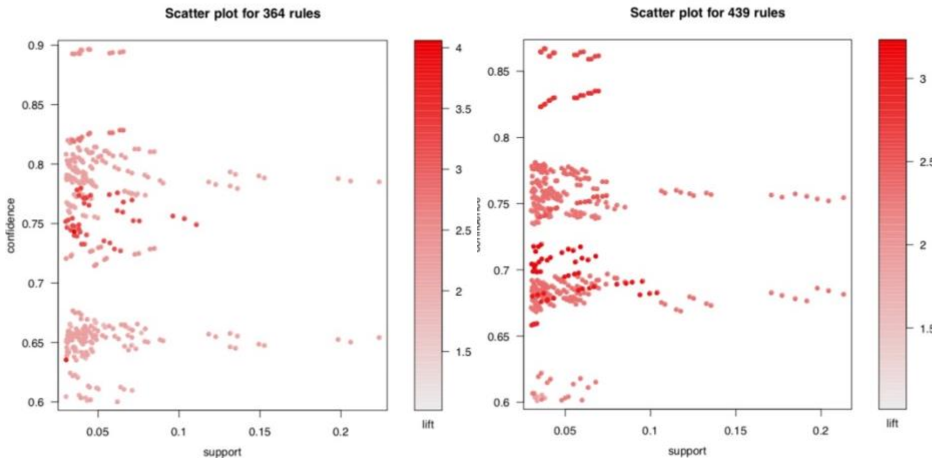
Once association rules have been obtained for each dataset separately, several graphics with interesting results are obtained. These graphics allow the results of the datasets to be analysed in a general way and the behaviour of each one to be compared. For example, in Fig. 4, the scatter plot of the four datasets displays values for support and confidence measures on the x-axis and y-axis, respectively. Additionally, the lift measure is represented on the right of each plot by the colour-coded points, higher values of the lift measure being darker. The total number of rules for each dataset is detailed in the upper part of the graph.

As can be observed, similar numbers of rules are obtained in spite of the difference in the input number of accidents in the datasets. At first glance, the behaviour of the rules seems generally similar in all phases. Most rules present a confidence between 65% and 85%, although some rules achieve a result above 85%. The support measure is similar in the four phases and the values are usual for these kind of problems with many variables. Similarly, the support value is regular, and, in this case, the values obtained are lower than 10%. Nevertheless, if we analyse them in more detail, some differences can be identified. For example, phases 2 and 3 present a more noticeable difference in terms of lift measurement, where the rules achieve a higher value, between 3.5 and 4. Remember that, as explained in Section 4, the higher the value of the lift measure is, the stronger is the association. So, from the values of all the measures, it can be considered that the rules obtained are satisfactory.



(a) Dataset of Phase 1

(b) Dataset of Phase 2



(c) Dataset of Phase 3

(d) Dataset of Phase 4

Fig. 4. Scatter plot for each dataset.

There are no established values to indicate whether the measures obtained for the rules are correct or incorrect. This is because these values depend on the characteristics of the dataset. However, it is known by the research community that the confidence should exceed 60% and the lift should exceed 1. However, the support is more dependent on the dataset.

Antecedents		Pred.	Phase 1					Phase 2					Phase 3					Phase 4				
Subcontrac	Subcontractor	17	X	X	X		X	X	X	X		X	X	X	X	X	X	X	X	X		X
R1	Risk Evaluation Yes	16	X	X	X		X	X	X	X		X	X	X	X		X	X	X	X		X
H1	Habitual work Yes	15	X	X	X		X	X	X	X		X		X	X		X	X	X	X		X
R2	Risk Evaluation No	13		X	X		X	X	X	X		X		X	X		X		X	X		X
S1	1-9 employees	11		X			X		X	X		X		X	X		X		X	X		X
A3	31-40 years	11		X	X		X		X	X		X		X	X		X		X			X
DW1	Monday	9		X			X		X			X		X	X		X		X			X
A2	21-30 years	9		X			X		X			X		X			X		X	X		X
S2	10-49 employees	8		X			X		X			X		X			X		X			X
S3	50-249 employees	8		X			X		X			X		X			X		X			X
DW2	Tuesday	8		X			X		X			X		X			X		X			X
DW3	Wednesday	8		X			X		X			X		X			X		X			X
DW4	Thursday	8		X			X		X			X		X			X		X			X
DW5	Friday	8		X			X		X			X		X			X		X			X
PA4	See Table 2	8		X			X		X			X		X			X		X			X
PA5	See Table 2	8		X			X		X			X		X			X		X			X
PA6	See Table 2	8			X		X			X		X			X		X			X		X
D5	See Table 2	8	X		X			X		X			X		X			X		X		
A4	41-50 years	8		X			X		X			X		X			X		X			X
LS3	2-6 months	8		X			X		X			X		X			X		X			X
LS5	1-3 years	8		X			X		X			X		X			X		X			X
LS4	6 months-1 year	6		X			X		X			X							X			X
PA2	See Table 2	5					X					X					X		X			X
M. Contrac	Main Contractor	4		X			X		X			X										
D7	See Table 2	4					X					X					X					X
FC3	See Table 2	4	X					X					X					X				
FC7	See Table 2	4		X					X					X					X			
S4	250-499 employees	2		X			X															
S5	500 employees or more	2		X			X															
PA3	See Table 2	2														X					X	
A5	51-60 years	2												X			X					
LS7	5-10 years	2												X			X					
Consequents			PA6	D7	FC3	FC4	FC7	PA6	D7	FC3	FC4	FC7	PA6	D7	FC3	FC4	FC7	PA6	D7	FC3	FC4	FC7

Table 5 Predominance of the category of variables in the association rules.
Note: The text length of some values was too long. Full versions can be seen in Table 2.

For the sake of simplicity and prior to the association analysis, to obtain more significant association rules, we have determined the predominance of the category of variables in the association rules and found that the occurrence of the variable “outsourcing” and “having the risks evaluated” were the main contributors, as can be seen in Table 5, where the comparison of these results by phases is more intuitive. On the left side, all the antecedents that appear in the rules are summarised together with the number of times they appear in each phase for each consequent (Pred). The antecedents represent the values of the variables that most frequently contribute to the substantiation of the work accident in the associations found. Using this illustration, the most common antecedents are quickly identified. In the middle part, the antecedents that emerge for each phase are marked with an *X*. Finally, in the lower part the consequent is represented. Therefore, summarised in this way, both the differences and the similarities between the different phases can be quickly observed.

The association rules obtained show five consequents that are involved in accidents in all phases of the work. These are:

• Physical Activity (PA6): movement

• Deviation (D7): Body movement under or with physical stress (generally leading to an internal injury)

• Contact (FC3): Horizontal or vertical impact with or against a stationary object (the victim is in motion)

• Contact (FC4): Struck by object in motion, collision with

• Contact (FC7): Physical overexertion or mental stress

It can be observed that these five consequents are grouped into three; the physical activity that the victim is engaged in at the time of the accident (described as being in motion); the abnormal event that originates the accident (described as movement of the body as a consequence of or with physical effort); and the way in which the victim has been injured, which involves three possibilities: hit against a stationary object, worker in motion, collision or blow against a moving object, and overexertion, mental trauma, radiation, noise, etc.

As can be seen in the Table, the attribute that has the highest frequency considering all the phases grouped together is “Subcontract”, that is, the company of the injured worker performing tasks on behalf of another company. This condition appears 17 out of 20 possible times. The attribute with the next highest number of occurrences (16 of 20) is “R1”, i.e. there is a risk assessment. Next are the following attributes whose level of relevance is due to their appearance in the association rules: it is a regular job (Habitual Work), there is no risk assessment (Risk Evaluation), the size of the company is from 1 to 9 full-time equivalent employees (Size of enterprise) and workers are aged between 30 and 40 years (Age). The remaining attributes do not appear in more than 50% of the possible cases, that is, they appear less than 10 times out of the 20 considered.

If each of the work phases is studied separately, it can be observed that in different phases, the consequents, antecedents and association rules have different peculiarities. Significant rules have been extracted based on their support (exceeding 10%), confidence (exceeding 75%) and lift (exceeding 2) values.

The results for each of the work phases are presented separately in the following subsections.

5.1. Foundation and structure phase

This phase corresponds to the beginning of the construction work. As can be observed in Table 5, there are two consequents that consider practically all the antecedents considered in this study. These two consequents refer to; (i) the abnormal event that originates the accident (deviation – D7), which is a movement of the body as a consequence of or with physical effort, and (ii) the way in which the victim has been injured (form of contact – FC7), which is an overexertion, mental trauma, radiation, noise, etc.

It is noteworthy that there are antecedents that are only relevant in this phase of the construction work. This is the case of S4 and S5, which show the influence of the number of workers working in the local unit of the workplace (from 250 to 499 employees and more than 500 full-time equivalent employees, respectively). These attributes are only part of the rules for the initial phase in a construction site, the foundation and structure phase.

Antecedents					Consequents	Support	Confidence	Lift
D7				O4	FC7	0.12	0.77	2.47
D7			R1		FC7	0.14	0.77	2.39
D7	H1				FC7	0.21	0.77	2.39
D7		OS			FC7	0.17	0.77	2.38
D7	H1			O4	FC7	0.12	0.77	2.48
D7	H1		R1		FC7	0.14	0.77	2.40
D7		OS	R1		FC7	0.11	0.77	2.40
D7	H1	OS			FC7	0.17	0.77	2.39
D7	H1	OS	R1		FC7	0.11	0.77	2.40

Table 6 Significant association rules in foundation and structure phase.

Regarding the association rules generated for the foundation/structure phase, nine rules that exceed the minima set for support, trust and lift measures have been found. It can be observed that their consequent is unique (FC7), which indicates the way in which the victim has been injured (form of contact). Concretely, it is overexertion, mental trauma, radiation, noise, etc. If we enumerate the antecedent of the association rules, in the nine rules the deviation D7 appears, which corresponds with a movement of the body as a consequence of or with physical effort. This is followed in importance by H1, which indicates that the work is performed in a habitual way. The antecedents OS – *Subcontract*- and S1 – *there is a risk assessment* – are also present. The rule with the most attributes indicates that there is a need to attend to the prevention of accidents caused by overexertion in subcontracted workers making movements of the body requiring physical effort, who are carrying out their usual work, even though there is a risk assessment, and particularly in the foundation and structure phase.

5.2. Facilities and masonry phase

This phase is carried out after the foundation/structure phase and it requires a long period of work on the construction site. As can be seen in Table 5, similar to the previous phase, two consequents D7 and FC7 are linked to most of the antecedents that appear in the rules obtained, although this number of antecedents is lower than the first phase. Among the predominant antecedents in the rules, there are two that appear with increased frequency in this phase. These attributes are: (i) related to the absence of risk assessment in this phase of installations/masonry (R2) and (ii) the number of workers working in the workplace or a company size between 1 and 9 full-time employees at this phase of the construction process.

Antecedents					Consequents	Support	Confidence	Lift
D7			R1		FC7	0.14	0.79	2.26
D7		O5			FC7	0.19	0.78	2.25
D7				OS	FC7	0.20	0.78	2.25
D7	H1				FC7	0.23	0.78	2.24
D7		O5	R1		FC7	0.11	0.79	2.26
D7	H1		R1		FC7	0.14	0.79	2.26
D7		O5		OS	FC7	0.16	0.79	2.25
D7	H1	O5			FC7	0.18	0.78	2.25
D7	H1			OS	FC7	0.20	0.78	2.25
D7	H1		R1	OS	FC7	0.12	0.79	2.26
D7	H1	O5	R1		FC7	0.11	0.79	2.27
D7	H1	O5		OS	FC7	0.15	0.79	2.26

Table 7 Significant association rules in facilities and masonry phase.

Concerning the association rules for the installations/masonry phase, twelve rules that meet the requirements for support, trust and lift measures are obtained. Similarly to the previous phase, its unique consequent is FC7. Regarding antecedents, in the twelve rules there appears a deviation concerning a movement of the body as a consequence of or with physical effort (D7). The second most frequent antecedent in the significant association rules indicates that the work performed by the worker was habitual work (H1), which appears in seven of the twelve rules. For the first time, an antecedent related to the occupation of the workers appears, which indicates that masonry workers are part of the consolidated accidental patterns while installers are not. Subcontractors (OS) and the existence of risk assessment (R1) are the other attributes that appear in the rules of this phase.

5.3. Completion phase

This phase can be identified as the completion phase on the construction site where painters, carpenters and tilers are involved. As can be seen in Table 5, again, the two consequents D7 and FC7 are affected by a greater number of antecedents.

In this phase, the antecedent *subcontractor* affects all the consequents. At the same time, *contractor* is eliminated as an influential antecedent for this phase of the work.

There are some antecedents that are relevant only in this completion phase, in this case, driving or being on board a means of transport or loading equipment (PA3). This attribute is also relevant for managers/labourers. Another antecedent relating to the age of the worker appears in the rules of this phase of work and exposes the influence of the age of the victim at the time of the accident, specifically the range between 50 and 60 years (A5). The rules add another influential value of the length of service in the work at the time of the accident, namely the range between 5 and 10 years (LS7).

Antecedents					Consequents	S	C	L
D7			R1		FC7	0.15	0.79	2.22
D7	OS				FC7	0.21	0.79	2.21
D7		H1			FC7	0.22	0.79	2.21
D7				O14	FC7	0.14	0.78	2.19
D5	OS				FC3	0.10	0.75	3.27
D7	OS		R1		FC7	0.13	0.79	2.23
D7		H1	R1		FC7	0.15	0.79	2.22
D7	OS	H1			FC7	0.19	0.79	2.22
D7	OS			O14	FC7	0.12	0.78	2.20
D7		H1		O14	FC7	0.13	0.78	2.20
D7	OS	H1		O14	FC7	0.12	0.78	2.21
D7	OS	H1	R1		FC7	0.13	0.79	2.23

Table 8 Significant association rules in completion phase.

In this phase, twelve association rules have been obtained that exceed the set parameters. In this phase a new consequent is added, related to the way in which the victim has been injured (FC3). Specifically, it is a blow against an immobile object by a worker in motion. In relation to the antecedents, the D7 deviation appears as in the previous phases. However, in the association rule that has FC3 as a consequent, the abnormal event that causes the accident or deviation is D5, that is, a fall or a slip or trip with a fall.

5.4. During the entire process

In this last section, all the professionals who perform their tasks during all phases of the construction work are considered (managers and labourers). In this case, we can observe again that the consequents D7 and FC7 are those that are affected by a greater number of antecedents.

Regarding the influence of the age of the victim at the time of the accident, there is an increase in the predominance of A2 (between 20 and 30 years), which appears in the rules of three of the five consequents.

The antecedent PA3 (driving or being on board a means of transport or loading equipment) is also relevant in this phase, as well as for the completion phase.

					Consequents	Support	Confidence	Lift
D7			R1		FC7	0.14	0.76	2.33
D7		OS			FC7	0.19	0.76	2.33
D7	H1				FC7	0.21	0.75	2.32
D7				O16	FC7	0.20	0.75	2.32
D7		OS	R1		FC7	0.12	0.76	2.33
D7	H1		R1		FC7	0.13	0.76	2.33
D7	H1	OS			FC7	0.18	0.76	2.33
D7			R1	O16	FC7	0.13	0.76	2.33
D7		OS		O16	FC7	0.18	0.75	2.33
D7	H1			O16	FC7	0.29	0.75	2.32
D7	H1	OS	R1		FC7	0.11	0.76	2.34
D7		OS	R1	O16	FC7	0.11	0.76	2.34
D7	H1		R1	O16	FC7	0.12	0.76	2.33
D7	H1	OS		O16	FC7	0.17	0.76	2.33
D7	H1	OS	R1	O16	FC7	0.11	0.76	2.34

Table 9 Significant association rules in occupations that work in all phases.

A total of fifteen association rules have been obtained for this phase that meet the requirements set for the measures considered. The unique consequent is FC7. In the antecedents, D7 appears in all the rules; that is, the deviation that causes the accident is a movement of the body as a consequence of or with physical effort. The rest of the relevant antecedents (there is a risk assessment, subcontract, habitual work, construction labourer) appear in eight of the fifteen rules, varying from those formed by two attributes to one formed by five. For a second time, a value of the occupation variable (construction labourer) appears in the significant association rules.

6. Discussion

The data used to achieve the objective of this study contains information on a total of 1,525,865 work accidents in the construction sector that occurred in Spain between 2003 and 2015. These data were obtained following the coding classification systems harmonised at the European level, and they present 11 variables per accident. The construction industry is diverse and globally a sector with comparatively high accident rates. The results of existing research on accidents in the construction industry show that the factors that influence the incidence of accidents are similar in many different countries (Gibb et al., 2014; Gürcanli & Müngen, 2009; Kartam et al., 2000; Siu et al., 2004; Guo & Goh, 2017) and that the reasons for the accidents and therefore the reduction of risk in construction works are complex and multifaceted (Hide et al., 2003; Haslam et al., 2005; Gambatese et al., 2008). In accordance with the general objective of this work, to highlight the relevance of various categories of variables reflected in the accident reports of the construction industry, the factor of subcontracting shows the highest level of importance. To this result we can add a behavior contrary to that of an important accident reduction factor in the risk assessment variable. Furthermore, the results show that the association rules of the variables differ according to the phase of the construction work. In the accident patterns of each phase, different variables are identified that may allow a more accurate evaluation of the real risks. These findings have practical implications for improving risk management, because they can allow various stakeholders in the construction industry, especially OHS professionals, to adjust the work organisation, improve decisions about preventive measures or action plans and, by extension, make prevention more effective. The results displayed as rules are easy to understand and can even serve to educate workers about occupational hazards by reading into an association rule a real situation that ended in an accident at the workplace.

According to the results, the variable of being a subcontracted worker appears in all phases of the works and in all association rules, except for the consequent (FC4) (Code Form _ Contact. Shock or blow against a moving object, collision with ...). For a long time, the implication of outsourcing in work accidents has generated a high interest on the part of researchers (Johnstone et al., 2000; Mayhew et al., 2001; Quinlan et al., 1997), for example, in the manufacturing industry (Nenonen, 2011), the petrochemical industry (Rebitzer, 1995), the clothing industry (Quinlan & Mayhew, 1999) or the mining industry (Blank et al., 1995). In the construction industry, there is evidence that the intensification of outsourcing has contributed in part to the problem of safety in construction (Debrah & Ofori, 2001). Outsourcing has also revealed an increase in the difficulty of site management by the prime contractor, and thus a negative effect on safety (Yung, 2009). Another result confirmed that the risk of accidents is higher for workers of subcontractors than for workers of the main contractor (Salminen, 1995) and it is a proven fact that companies that make excessive use of subcontractors have more accidents (Kartam et al., 2000). The highest accident rate corresponding to workers from subcontracted companies in the construction sector was in the period 2011–2013 in Spain, when it was 2.5 times higher than the average of other sectors (INSHT, 2016).

Clearly, the effect of outsourcing on workplace accidents in the construction industry has been addressed for decades and in various ways. The fact that this variable is revealed in this study implies that there is still much work to be done in this field. Despite exhaustive investigation of the accident rate linked to subcontracting in the construction industry, there is still a failure to find adequate preventive strategies, to which the difficulty of having accurate subcontracting data for investigation contributes.

Another of the outstanding results in the search for association rules in accident events in the construction industry has been the risk assessment variable. The literature provides abundant evidence of the importance of carrying out risk assessment in this sector (Fung et al., 2010; Romero & Gámez, 2005) and of its key role in the management of occupational risks (Reddy, 2015; Sousa et al., 2014; Wehbe et al., 2016). However, several investigations have called into question many aspects of risk assessment. There are studies that show the difficulty in practice of risk assessment in the construction industry (Santiago, 2010). Other studies determine the existence of deficiencies in current general practice for risk assessment (Fung et al., 2012). Research shows that there is no systematic approach to risk assessment, but rather that risks are assessed based on individual judgment driven by experience and educational background (Fung et al., 2010). Other research confirms the existence of a gap between the theory and practice of risk assessment in the construction industry (Taroun, 2014). There are results that demonstrate a concentration of factors underlying fatal construction accidents associated with deficiencies in planning and risk assessment (Hale et al., 2012). And other findings highlight that the repercussions of deficiencies in risk assessment in the construction sector are serious, especially for workers (Cantalejo et al., 2005).

In the present investigation we have found that both the existence and the absence of an occupational risk assessment are factors found in the association rules identified. This suggests that merely conducting a risk assessment is not effective in fulfilling its role in preventing workplace accidents. Our results corroborate findings that show that risk assessments, especially serious in the construction sector, are considered a procedure, lack precision and are not effective enough to control real risks (Swuste et al., 2012). It is essential to move towards a more specific risk assessment methodology in construction, adapted to the particular circumstances of this industry. A correct risk assessment and proposed preventive measures or action plans are considered prominent elements of an effective safety system, but they must be established using experience, historical data and numerical methods (Bilir & Gürcanli, 2018). For risk management to be effective, it must be based on consensus and have the collaboration of all involved, although such integration is difficult to achieve.

Regarding one of the specific objectives of this study to analyze whether the association rules that express relationships between various variables in occupational accidents in construction sites are different in each construction phase, the results shown in Tables 6 to 9 justify compliance with this theory. Studies have shown that accident patterns are different for different types of construction projects (Cheng et al., 2012; Shao et al., 2019) or for different construction works (Bilir & Gürcanli, 2018). This fact also occurs in the different phases of construction works. In different phases, variables appear in the association rules that are exclusive to the works that are executed in each phase and the conditions in which they are developed. These results coincide with those obtained in other studies on the significance of the construction phase for the way an accident occurs (INSHT, 2011; INSHT, 2016) and in previous investigations on accident mechanisms in different phases of a construction project (Carrillo et al., 2017). Partitioning the project into several parts to make activities more manageable and determining work items accurately are beneficial efforts for effective activity-based risk assessment (Gürcanli et al., 2015). A different approach is required to identify hazards and assess risks, increase safety and prevent accidents, especially in construction (Rozenfeld et al., 2010).

This research shows that the age or length of service ranges present different results in different phases of a construction. For example, it is only in the phase that could be identified as finishes that the age range of the victim at the time of the accident appears as a variable to consider. This result could serve as a bridge between the results obtained by authors who found the relevance of age ranges of workers in the construction industry, for example by finding that there is a high correlation between age and other personal factors of the injured workers (Amiri et al., 2016; Liao and Perng, 2008) or that the main consolidated rule indicates that the possibility of accidents is highest in the age range of 45 to 54 (Liao and Perng, 2008), and the findings obtained by other authors who state that accident rates are not related to age (Siu et al., 2003), or that age does not appear as a condition in the rules of occurrences of public or private projects (Cheng et al., 2012), since in all these cases, the accident analysis was carried out on construction works in general, without dividing their execution into phases. There are obvious differences in the variable of the age of workers and accident patterns in the construction industry. Findings on the underlying causes of these disparities, as well as other variables such as company size or occupation, should be investigated further (Shin et al., 2018).

7. Conclusion

The purpose of this study is to investigate the occupational accidents reported at construction sites in Spain for more than a decade and to highlight predominant attributes and accident patterns, not in an isolated way but as a real situation in the context of construction activities. The results show that there are commonly recurring factors with multiple relationships and in all phases of works. They also show specific factors and patterns at different phases of construction. Based on these findings, some useful recommendations in the field of occupational risk prevention on construction sites are presented. Therefore, the values and rules found in this document can provide a valuable reference to all agents involved in the construction industry (employers, regulators, inspectors, professionals, workers) to improve risk management, preventive measures and action plans.

The results reveal that subcontracting in the construction industry is a critical factor in occupational accidents. It has been shown that in all construction phases it is a relevant variable. One recommendation would be to increase efforts to convince companies that, when selecting a subcontracted company, it is necessary to consider aspects of that company related to occupational health and safety, such as its accident rate or its OHS certifications. On the other hand, requirements should be increased for small companies and self-employed workers that are outsourced, to ensure their education and training in OHS matters. It would be advisable for legislators to put the necessary limits on the practice of subcontracting, helping to eliminate the integration and coordination problems that contribute to accidents. This should also improve the collection of information on this section to facilitate accident investigation. We have run into the limitation due to lack of information about the nearby activities of other subcontractors or workers that could cause accidents.

The results have also shown that risk assessment is not an insurmountable barrier to accidents. It is advisable to increase the adaptation of risk assessments to construction processes and their progress on the site of construction. This should help improve the basis for preventive decision-making after risk assessment.

Regarding the contributions to adapt risk assessments to construction processes and their evolution, it is possible to:

• Establish a permanent training system for evaluators with the aim of keeping up-to-date knowledge and skills on risk management methods, construction materials, the evolution of construction processes and preventive measures.

• Incorporate measures to control the variability of assessors due to individual factors and avoid predetermined assessments of the level of risk. This would facilitate the closing of the gap between the theory and practice of risk assessment.

It is interesting to analyse the accident patterns in each of the work phases separately. In the different phases of the works, different actions on working conditions must be promoted. This will allow adequate planning and prevention of occupational risks based on the knowledge of the accident patterns that occur in each phase.

The identification of accident patterns directly related to the phases of the construction site can provide valuable information for developing accident prevention strategies. This would be more effective than simply identifying the causal factors on a construction site. With this knowledge, the effectiveness of the measures or action plans to be implemented to reduce the accident rate can be increased.

Although our source of information for this study is in Spain, the fact that we have used the coding of the variables according to the "Accident Statistics Preparation Methodology of the European Commission (ESAW)" suggests conducting similar studies in other countries with identical coding considering the different phases of work.

The results of this work can be used as preliminary results to obtain a more accurate and complete representation of the genesis of accidents in each of the phases of the construction work. On the other hand, they can serve as a basis to deepen the relationship between the factors that have influenced the occurrence of accidents or the dependence between the causal factors.

The findings of this study can provide basic information for those responsible for managing occupational hazards in the construction industry. Consideration of the different stages of the construction work will improve their actions in terms of risk reduction and control. It is important to encourage these professionals to incorporate other types of factors in their technical decisions and to share experiences with other professionals.

Acknowledgements

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In ACM sigmod record (pp. 207–216). ACM volume 22.

Agrawal, R., Srikant, R. et al. (1994). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (pp. 487–499). volume 1215.

Al-Bayati, A. J., Abudayyeh, O., & Albert, A. (2018). Managing active cultural differences in U.S. construction workplaces: Perspectives from non-hispanic workers. Journal of Safety Research, 66, 1 – 8. URL: http://www.sciencedirect.com/science/article/pii/S002243751730765X. doi:https://doi.org/10.1016/j.jsr.2018.05.004.

Amiri, M., Ardeshir, A., Fazel Zarandi, M. H., & Soltanaghaei, E. (2016). Pattern extraction for high-risk accidents in the construction industry: a data-mining approach. International journal of injury control and safety promotion, 23(3), 264-276.

Arquillos, A. L., Rubio-Romero, J.C., & Gibb, A. (2012). Analysis of construction accidents in Spain, 2003-2008. Journal of Safety Research, 43,381 – 388. URL: http://www.sciencedirect.com/science/article/pii/S0022437512000795. doi:https://doi.org/10.1016/j.jsr.2012.07.005.

Bilir, S., & Gürcanli, G. E. (2018). A method for determination of accident probability in construction industry. Teknik Dergi, 29(4), 8537-8561. Doi: https://doi.org/10.18400/tekderg.363613

Blank, V. L., Andersson, R., Lindén, A., & Nilsson, B. C. (1995). Hidden accident rates and patterns in the Swedish mining industry due to involvement of contractor workers. Safety Science, 21(1), 23-35.

Cantalejo, A. F., Aibar, M. M. T., & de la Orden Rivera, M. V. (2005). Análisis de los accidentes de trabajo mortales en España. Prevención, trabajo y salud: Revista del Instituto Nacional de Seguridad e Higiene en el Trabajo, (34), 24-42.

Carrillo-Castrillo, J. A., Trillo-Cabello, A. F., & Rubio-Romero, J. C. (2017). Construction accidents: Identification of the main associations between causes, mechanisms and stages of the construction process. International journal of occupational safety and ergonomics, 23(2), 240-250.

Chau, N., Gauchard, G. C., Siegfried, C., Benamghar, L., Dangelzer, J.-L., Fran¸cais, M., Jacquin, R., Sourdot, A., Perrin, P. P., & Mur, J.-M. (2004). Relationships of job, age, and life conditions with the causes and severity of occupational injuries in construction workers. International Archives of

Occupational and Environmental Health, 77, 60–66. URL: https://doi.org/10.1007/s00420-003-0460-7. doi:10.1007/s00420-003-0460-7.

Cheng, C.-W., Leu, S.-S., Cheng, Y.-M., Wu, T.-C., & Lin, C.-C. (2012). Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. Accident Analysis & Prevention, 48, 214 – 222. URL: http://www.sciencedirect.com/science/article/pii/S0001457511000959. doi: https://doi.org/10.1016/j.aap.2011.04.014. Intelligent Speed Adaptation + Construction Projects.

Cheng, C.-W., Leu, S.-S., Lin, C.-C., & Fan, C. (2010a). Characteristic analysis of occupational accidents at small construction enterprises. Safety Science, 48, 698 – 707. URL: http://www.sciencedirect.com/science/article/pii/S0925753510000317. doi:https://doi.org/10.1016/j.ssci.2010.02.001.

Cheng, C.-W., Lin, C.-C., & Leu, S.-S. (2010b). Use of association rules to explore cause–effect relationships in occupational accidents in the taiwan construction industry. Safety Science, 48,436 – 444. URL: http://www.sciencedirect.com/science/article/pii/S0925753509002161. doi:https://doi.org/10.1016/j.ssci.2009.12.005.

Cheng, Y., der Yu, W., & Li, Q. (2015). Gâbased multi-level association rule mining approach for defect analysis in the construction industry. Automation in Construction, 51,78 – 91. URL: http://www.sciencedirect.com/science/article/pii/S092658051400260X. doi:https://doi.org/10.1016/j.autcon.2014.12.016.

Chi, C.-F., Yang, C.-C., & Chen, Z.-L. (2009). In-depth accident analysis of electrical fatalities in the construction industry. International Journal of Industrial Ergonomics, 39, 635 – 644. URL: http://www.sciencedirect.com/science/article/pii/S0169814108000073. doi:https://doi.org/10.1016/j.ergon.2007.12.003. Special issue: Felicitating Colin G. Drury.

De Solminihac, H. (2011). Procesos y técnicas de construcción. Ediciones UC.

Debrah, Y. A., & Ofori, G. (2001). Subcontracting, foreign workers and job safety in the Singapore construction industry. Asia Pacific Business Review, 8(1), 145-166.

Dong, C., Wang, F., Li, H., Ding, L., & Luo, H. (2018). Knowledge dynamicsintegrated map as a blueprint for system development: Applications to safety risk management in Wuhan metro project. Automation in Construction, 93, 112 – 122. URL: http://www.sciencedirect.com/science/article/pii/S0926580517309561.doi:https://doi.org/10.1016/j.autcon.2018.05.014.

Eastman, C., min Lee, J., suk Jeong, Y., & kook Lee, J. (2009). Automatic rule-based checking of building designs. Automation in Construction, 18, 1011 – 1033.URL:http://www.sciencedirect.com/science/article/pii/S0926580509001198. doi:https://doi.org/10.1016/j.autcon.2009.07.002.

del Estado (BOE)., B. O. (). Orden tas/2926/2002, de 19 de noviembre, por la que se establecen nuevos modelos para la notificación de los accidentes de trabajo y se posibilita su transmisión por procedimiento electrónico. Spanish government; 2002. EuropeanStatisticalOffice (2019). URL: https://ec.europa.eu/eurostat.

Eurostat (2013). European Statistics on Accidents at Work. Methodology-2001 Edition. doi:10.2785/40882.

Fernández-Muñiz, B., Montes-Peón, J. M., & Vázquez-Ordás, C. J. (2009). Relation between occupational safety management and firm performance. Safety Science, 47, 980 – 991. URL: http://www.sciencedirect.com/science/article/pii/S0925753508001951.doi:https://doi.org/10.1016/j.ssci.2008.10.022.

Fung, I. W., Tam, V. W., Lo, T. Y., & Lu, L. L. (2010). Developing a risk assessment model for construction safety. International Journal of Project Management, 28(6), 593-600.

Fung, I. W., Tam, V. W., Lo, T. Y., & Lu, L. L. (2010). Developing a risk assessment model for construction safety. International Journal of Project Management, 28(6), 593-600.

Fung, I. W., Lo, T. Y., & Tung, K. C. (2012). Towards a better reliability of risk assessment: Development of a qualitative & quantitative risk evaluation model (Q2REM) for different trades of construction works in Hong Kong. Accident Analysis & Prevention, 48, 167-184.

Gambatese, J. A., Behm, M., & Rajendran, S. (2008). Design's role in construction accident causality and prevention: Perspectives from an expert panel. Safety science, 46(4), 675-691.

Gao, R., Chan, A. P., Utama, W. P., & Zahoor, H. (2016). Workers' perceptions of safety climate in international construction projects: Effects of nationality, religious belief, and employment mode. Journal of Construction Engineering and Management, 143, 04016117.

Gibb, A., Lingard, H., Behm, M., & Cooke, T. (2014). Construction accident causality: learning from different countries and differing consequences. Construction Management and Economics, 32(5), 446-459.

Golizadeh, H., Hon, C. K., Drogemuller, R., & Hosseini, M. R. (2018). Digital engineering potential in addressing causes of construction accidents. Automation in Construction, 95, 284-295.

Gürcanli, G. E., & Müngen, U. (2009). An occupational safety risk analysis method at construction sites using fuzzy sets. International Journal of Industrial Ergonomics, 39(2), 371-387.

Gunduz, M., Birgonul, M. T., & Ozdemir, M. (2017). Fuzzy structural equation model to assess construction site safety performance. Journal of Construction Engineering and Management, 143(4), 04016112.

Guo, B. H., & Goh, Y. M. (2017). Ontology for design of active fall protection systems. Automation in Construction, 82, 138-153. Doi: https://doi.org/10.1016/j.autcon.2017.02.009

Guo, S., Zhang, P., & Ding, L. (2019). Time-statistical laws of workers'a unsafe behavior in the construction industry: A case study. Physica A: Statistical Mechanics and its Applications, 515, 419 – 429. URL: http://www.sciencedirect.com/science/article/pii/ S0378437118312238. doi:https://doi.org/10.1016/j.physa.2018.09.091.

Gürcanli, G. E., Bilir, S., & Sevim, M. (2015). Activity based risk assessment and safety cost estimation for residential building construction projects. Safety Science, 80, 1-12. doi: https://doi.org/10.1016/j.ssci.2015.07.002

Hale, A., Walker, D., Walters, N., & Bolt, H. (2012). Developing the understanding of underlying causes of construction fatal accidents. Safety science, 50(10), 2020-2027.

Hasebe, T., & Sakai, T. (2018). Are elderly workers more likely to die in occupational accidents? evidence from both industry-aggregated data and administrative individual-level data in japan. Japan and the World Economy, 48,79-89.

URL:http://www.sciencedirect.com/science/article/pii/S0922142518300483. doi:https://doi.org/10.1016/j.japwor.2018.09.001.

Haslam, R. A., Hide, S. A., Gibb, A. G., Gyi, D. E., Pavitt, T., Atkinson, S., & Duff, A. R. (2005). Contributing factors in construction accidents. Applied ergonomics, 36(4), 401-415.

Heinrich, H. W. (1941). Industrial Accident Prevention. A Scientific Approach. Industrial Accident Prevention. A Scientific Approach., (Second Edition).

Hide, S., Atkinson, S., Pavitt, T. C., Haslam, R., Gibb, A. G., & Gyi, D. E. (2003). Causal factors in construction accidents.

Hola, B., & Szóstak, M. (2017a). Methodology of analysing the accident rate in the construction industry. Procedia Engineering, 172, 355 – 362. URL: http://www.sciencedirect.com/science/article/pii/S1877705817305465. doi:https://doi.org/10.1016/j.proeng.2017.02.040. Modern Building Materials, Structures and Techniques.

Hola, B., & Szóstak, M. (2017b). An occupational profile of people injured in accidents at work in the polish construction industry. Procedia Engineering, 208,43–51. URL: http://www.sciencedirect.com/science/article/pii/S1877705817360095. doi:https://doi.org/10.1016/j.proeng.2017.11.019. Innovative Solutions in Construction Engineering and Management. Flexibility in Sustainable Construction.

INSHT (2011). Análisis de mortalidad por accidente de trabajo en España. 2008-2009-2010. Instituto Nacional de Seguridad e Higiene en el Trabajo. Ministerio de Empleo y Seguridad. NIPO: 272-13-022-x. España.

INSHT (2016). Análisis de mortalidad por accidente de trabajo en España. 2011 – 2012 - 2013. Instituto Nacional de Seguridad e Higiene en el Trabajo. Ministerio de Empleo y Seguridad. NIPO: 272-15-093-7. España.

Jo, B., Lee, Y., Kim, J., & Khan, R. (2017). Trend analysis of construction industrial accidents in korea from 2011 to 2015. Sustainability, 9, 1297.

Johnstone, R., Mayhew, C., & Quinlan, M. (2000). Outsourcing risk-the regulation of occupational health and safety where subcontractors are employed. Comp. Lab. L. & Pol'y J., 22, 351.

Kale, Ö. A., & Baradan, S. (2020). Identifying factors that contribute to severity of construction injuries using logistic regression model. Teknik Dergi, 31(2), 9919-9940.

Kang, Y., Siddiqui, S., Suk, S. J., Chi, S., & Kim, C. (2017). Trends of fall accidents in the U.S. construction industry. Journal of Construction Engineering and Management, 143, 04017043.

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

Knime analytics platform., 2020. URL https://www.knime. com/.

Li, H., Li, X., Luo, X., & Siebert, J. (2017). Investigation of the causality patterns of non-helmet use behavior of construction workers. Automation in Construction, 80, 95 – 103. URL: http://www.sciencedirect. com/science/article/pii/S0926580517301541. doi:https://doi.org/10.1016/j.autcon.2017.02.006.

Liao, C.-W., & Perng, Y.-H. (2008). Data mining for occupational injuries in the taiwan construction industry. Safety Science, 46, 1091 – 1102. URL: http://www.sciencedirect.com/science/article/pii/S0925753507000574. doi:https://doi.org/10.1016/j.ssci.2007.04.007.

Liu, P. (2018). Association rule analysis of factors contributing to extraordinarily severe traffic crashes in china. Journal of Safety Research, 67, 65 – 75. URL: http://www.sciencedirect. com/science/article/pii/S0022437518300690. doi:https://doi.org/10.1016/j.jsr.2018.09.013.

Loosemore, M., & Malouf, N. (2019). Safety training and positive safety attitude formation in the australian construction industry. Safety Science, 113, 233 – 243. URL: http://www.sciencedirect. com/science/article/pii/S0925753517304988. doi:https://doi.org/10.1016/j.ssci.2018.11.029.

López, M. A. C., Ritzel, D. O., Fontaneda, I., & Alcantara, O. J. G. (2008). Construction industry accidents in spain. Journal of Safety Research, 39, 497-507.URL:http://www.sciencedirect.com/science/article/pii/S0022437508001229. doi:https://doi.org/10.1016/j.jsr.2008.07.006.

López-Arquillos, A., Rubio-Romero, J. C., & Gibb, A. (2015). Accident data study of concrete construction companies' similarities and differences between qualified and non-qualified workers in Spain. International journal of occupational safety and ergonomics, 21, 486–492.

Lozano-Díez, R. V., López-Zaldívar, O., del Cura, S. H., & Verdú Vázquez, A. (2019). Analysis of the impact of health and safety coordinator on construction site accidents: The case of Spain. Journal of Safety Research, 68, 149 – 156. URL: http://www.sciencedirect. com/science/article/pii/S0022437518304924. doi:https://doi.org/10.1016/j.jsr.2018.12.012.

Martínez-Aires, M. D., López-Alonso, M., & Martínez-Rojas, M. (2018).
Building information modeling and safety management: A systematic review. Safety Science, 101, 11 – 18. URL: http://www.sciencedirect. com/science/article/pii/S0925753517314340.
doi:https://doi.org/10.1016/j.ssci.2017.08.015.

Martínez-Rojas, M., Marín, N., & Vila, M. A. (2013, June). A preliminary approach to classify work descriptions in construction projects. In 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS) (pp. 1090-1095). IEEE.

Martínez -Rojas, M., Marín, N., Molina, C., & Vila, M. (2015). Cost analysis in construction projects using fuzzy OLAP cubes. In 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).

Martínez-Rojas, M., Soto-Hidalgo, J. M., Marín, N., & Vila, M. A. (2018). Using Classification Techniques for Assigning Work Descriptions to Task Groups on the Basis of Construction Vocabulary. Computer-Aided Civil and Infrastructure Engineering, 33(11), 966-981.

Martínez-Rojas, M., Antolín, R. M., Salguero-Caparrós, F., & Rubio-Romero, J.C. (2020). Management of construction safety and health plans based on automated content analysis. Automation in Construction, 120,103362. Doi: https://doi.org/10.1016/j.autcon.2020.103362

Martínez-Rojas, M., Soto-Hidalgo, J. M., Martínez-Aires, M. D., & Rubio-Romero, J. C. (2021). An analysis of occupational accidents involving national and international construction workers in Spain using association rule technique. International journal of occupational safety and ergonomics, 1-37. Doi: https://doi.org/10.1080/10803548.2021.1901433

Marzouk, M., & Enaba, M. (2019). Text analytics to analyze and monitor construction project contract and correspondence. Automation in Construction, 98, 265 – 274. URL: http://www.sciencedirect.com/science/article/pii/S0926580517311330. doi:https://doi.org/10.1016/j.autcon.2018.11.018.

Mayhew, C., Quintan, M., & Ferris, R. (1997). The effects of subcontracting/outsourcing on occupational health and safety: survey evidence from four Australian industries. Safety Science, 25(1-3), 163-178.

Mínguez, J. B. P., & Moreno, A. S. (2004). Calidad del diseño en la construcción. Ediciones Díaz de Santos.

Ministerio de Trabajo, Migraciones y Seguridad Social (2019). Labour, migrations and social security Ministery. URL: http://www.mitramiss.gob.es/.

Mohammadi, A., Tavakolan, M., & Khosravi, Y. (2018). Factors influencing safety performance on construction projects: A review. Safety Science, 109, 382-397.
URL:http://www.sciencedirect.com/science/article/pii/S0925753516304234. doi:https://doi.org/10.1016/j.ssci.2018.06.017.

Nenonen, S. (2011). Fatal workplace accidents in outsourced operations in the manufacturing industry. Safety Science, 49(10), 1394-1403.

Newaz, M. T., Davis, P., Jefferies, M., & Pillay, M. (2019). Using a psychological contract of safety to predict safety climate on construction sites. Journal of Safety Research,68,9–19.URL: http://www.sciencedirect.com/science/article/pii/S0022437518304079. doi:https://doi.org/10.1016/j.jsr.2018.10.012.

Nielsen, K. J. (2014). Improving safety culture through the health and safety organization: A case study. Journal of Safety Research,48, 7 – 17. URL: http://www.sciencedirect.com/science/article/pii/S0022437513001552.doi:https://doi.org/10.1016/j.jsr.2013.10.003.

Pardo-Ferreira, M. D. C., Martínez-Rojas, M., Salguero-Caparrós, F., & Rubio-Romero, J. C. (2019). Evolution of the Functional Resonance Analysis Method (FRAM) through the combination with other methods. Dirección y Organización, 41-50

Poh, C. Q., Ubeynarayana, C. U., & Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. Automation in Construction, 93, 375 – 386. URL: http://www.sciencedirect.com/science/article/pii/S0926580517309147.

doi:https://doi.org/10.1016/j.autcon.2018.03.022.

Quinlan, M., & Mayhew, C. (1999). The effects of outsourcing on occupational health and safety: a comparative study of factory-based workers and outworkers in the Australian clothing industry. International journal of health services, 29(1), 83-107.

Quinlan, M., Mayhew, C., & Bohle, P. (2001). The global expansion of precarious employment, work disorganization, and consequences for occupational health: a review of recent research. International journal of health services, 31(2), 335-414.

Rameezdeen, R., & Elmualim, A. (2017). The impact of heat waves on occurrence and severity of construction accidents. International journal of environmental research and public health, 14, 70.

Reason, J. (2016). Managing the risks of organizational accidents. Routledge.

Rebitzer, J. B. (1995). Job safety and contract workers in the petrochemical industry. Industrial Relations: A Journal of Economy and Society, 34(1), 40-57.

Reddy, A. S. (2015). Risk management in construction industry-a case study. International Journal of Innovative Research in Science, Engineering and Technology, 4(10).

Rivas, T., Paz, M., Martín, J., Matías, J., García, J., & Taboada, J. (2011). Explaining and predicting workplace accidents using datamining techniques. Reliability Engineering & System Safety, 96,739 – 747. URL: http://www.sciencedirect.com/science/article/pii/S0951832011000342. doi:https://doi.org/10.1016/j.ress.2011.03.006.

Rozenfeld, O., Sacks, R., Rosenfeld, Y., & Baum, H. (2010). Construction Job Safety Analysis. Safety Science, 48(4), 491–498. Doi: https://doi.org/10.1016/j.ssci.2009.12.017

Rubio-Romero, J. C., & Gámez, M. D. C. R. (2005). Manual de coordinación de seguridad y salud en las obras de construcción. Ediciones Díaz de Santos.

Salguero-Caparros, F., Suarez-Cebador, M., & Rubio-Romero, J. C. (2015). Analysis of investigation reports on occupational accidents. Safety science, 72, 329-336.

Salminen, S. (1995). Serious occupational accidents in the construction industry. Construction Management and Economics, 13(4), 299-306.

Sanmiquel, L., Rossell, J. M., & Vintró, C. (2015). Study of Spanish mining accidents using data mining techniques. Safety Science, 75, 49 – 55. URL: http://www.sciencedirect.com/science/article/pii/S092575351500017X. doi:https://doi.org/10.1016/j.ssci.2015.01.016.

Santiago, J. A. E. (2010). Coordinadores de seguridad y salud en el sector de la construcción. Lex Nova.

Shao, B., Hu, Z., Liu, Q., Chen, S., & He, W. (2019). Fatal accident patterns of building construction activities in china. Safety Science, 111,253 – 263. URL: http://www.sciencedirect.com/science/article/pii/S0925753517315394. doi:https://doi.org/10.1016/j.ssci.2018.07.019.

Shepherd, R., Lorente, L., Vignoli, M., Nielsen, K., & Peiró, J. M. (2021). Challenges influencing the safety of migrant workers in the construction industry: A qualitative study in Italy, Spain, and the UK. Safety Science, 142, 105388. Doi:_ https://doi.org/10.1016/j.ssci.2021.105388

Shin, D.-P., Park, Y.-J., Seo, J., & Lee, D.-E. (2018). Association rules mined from construction accident data. KSCE Journal of Civil Engineering,22, 1027–1039. URL: https://doi.org/10.1007/s12205-017-0537-6. doi:10.1007/s12205-017-0537-6.

Silva, J. F., & Jacinto, C. (2012). Finding occupational accident patterns in the extractive industry using a systematic data mining approach. Reliability Engineering & System Safety, 108, 108– 122. URL: http://www.sciencedirect.com/science/article/pii/ S0951832012001354. doi:https://doi.org/10.1016/j.ress.2012.07.001.

Siu, O. L., Phillips, D. R., & Leung, T. W. (2003). Age differences in safety attitudes and safety performance in Hong Kong construction workers. Journal of Safety Research, 34(2), 199-205.

Siu, O. L., Phillips, D. R., & Leung, T. W. (2004). Safety climate and safety performance among construction workers in Hong Kong: The role of psychological strains as mediators. Accident Analysis & Prevention, 36(3), 359-366.

Sousa, V., Almeida, N. M., & Dias, L. A. (2014). Risk-based management of occupational safety and health in the construction industry–Part 1: Background knowledge. Safety science, 66, 75-86.

Stoilkovska, B. B., Zîleska Pancovska, V., & Mijoski, G. (2015). Relationship of safety climate perceptions and job satisfaction among employees in the construction industry: the moderating role of age. International journal of occupational safety and ergonomics, 21, 440–447.

Swuste, P., Frijters, A., & Guldenmund, F. (2012). Is it possible to influence safety in the building sector? : A literature review extending from 1980 until the present. Safety science, 50(5), 1333-1343.

Taroun, A. (2014). Towards a better modelling and assessment of construction risk: Insights from a literature review. International journal of Project management, 32(1), 101-115.

Trillo-Cabello, A. F., Carrillo-Castrillo, J. A., & Rubio-Romero, J. C. (2021). Perception of risk in construction. Exploring the factors that influence experts in occupational health and safety. Safety Science, 133, 104990.

Umer, W., Li, H., Lu, W., Szeto, G. P. Y., & Wong, A. Y. (2018). Development of a tool to monitor static balance of construction workers for proactive fall safety management. Automation in Construction, 94, 438-448. Doi: https://doi.org/10.1016/j.autcon.2018.07.024

Valle, M. A., Ruz, G. A., & Morrás, R. (2018). Market basket analysis: Complementing association rules with minimum spanning trees. Expert Systems with Applications, 97, 146 – 162. URL: http://www.sciencedirect.com/science/article/pii/S0957417417308503. doi:https://doi.org/10.1016/j.eswa.2017.12.028.

Wang, X., Huang, X., Luo, Y., Pei, J., & Xu, M. (2018). Improving workplace hazard identification performance using data mining. Journal of Construction Engineering and Management, 144, 04018068. doi:10.1061/ (ASCE)CO.1943-7862.0001505.

Wehbe, F., Al Hattab, M., & Hamzeh, F. (2016). Exploring associations between resilience and construction safety performance in safety networks. Safety science, 82, 338-351.

Winge, S., & Albrechtsen, E. (2018). Accident types and barrier failures in the construction industry. Safety science, 105, 158-166. Doi: https://doi.org/10.1016/j.ssci.2018.02.006

Winge, S., Albrechtsen, E., & Mostue, B. A. (2019). Causal factors and connections in construction accidents. Safety Science, 112130 – 141. URL: http://www.sciencedirect.com/science/article/pii/S0925753518309160. doi:https://doi.org/10.1016/j.ssci.201810.015.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Xiao, F., & Fan, C. (2014). Data mining in building automation system for improving building operational performance. Energy and Buildings, 75, 109 – 118. URL: http://www.sciencedirect. com/science/article/pii/S0378778814001169. doi:https://doi.org/10.1016/j.enbuild.2014.02.005.Xu, C., Bao, J., Wang, C., &

Xu, R., & Luo, F. (2021). Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest. Safety science, 135, 105125.

Yung, P. (2009). Institutional arrangements and construction safety in China: an empirical examination. Construction Management and Economics, 27(5), 439-450.

Zhang, F., Fleyeh, H., Wang, X., & Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. Automation in Construction, 99, 238 – 248. URL: http:// www.sciencedirect.com/science/article/pii/S0926580518306137. doi:https://doi.org/10.1016/j.autcon.2018.12.016.

Zhang, M., Shi, R., & Yang, Z. (2020). A critical review of vision-based occupational health and safety monitoring of construction site workers. Safety science, 126, 104658. Doi: https://doi.org/10.1016/j.ssci.2020.104658