Accident Analysis and Prevention 70 (2014) 155–166

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Extracting recurrent scenarios from narrative texts using a Bayesian network: Application to serious occupational accidents with movement disturbance

F. Abdat [a], S. Leclercq [a,*], X. Cuny [b], C. Tissot [c]

[a] INRS – Working Life Department, 1 rue de Morvan, 54500 Vandoeuvre les Nancy, France
[b] CNAM – Honorary Professor of Occupational Hygiene and Safety, 292 rue Saint-Martin, 75003 Paris, France
[c] INRS – Library & Literature Watch Division, 65 boulevard Richard Lenoir, 75011 Paris, France

ARTICLE INFO

Article history:
Received 21 May 2013
Received in revised form 2 April 2014
Accepted 7 April 2014
Available online 25 April 2014

Keywords:
Bayesian network
Recurrent scenarios
Narrative text
Occupational accident with movement disturbance

ABSTRACT

A probabilistic approach has been developed to extract recurrent serious Occupational Accident with Movement Disturbance (OAMD) scenarios from narrative texts within a prevention framework. Relevant data extracted from 143 accounts was initially coded as logical combinations of generic accident factors. A Bayesian Network (BN)-based model was then built for OAMDs using these data and expert knowledge. A data clustering process was subsequently performed to group the OAMDs into similar classes from generic factor occurrence and pattern standpoints. Finally, the Most Probable Explanation (MPE) was evaluated and identified as the associated recurrent scenario for each class. Using this approach, 8 scenarios were extracted to describe 143 OAMDs in the construction and metallurgy sectors. Their recurrent nature is discussed.

Probable generic factor combinations provide a fair representation of particularly serious OAMDs, as described in narrative texts. This work represents a real contribution to raising company awareness of the variety of circumstances, in which these accidents occur, to progressing in the prevention of such accidents and to developing an analysis framework dedicated to this kind of accident.

1. Introduction

Prevention of trips, collisions, slips and other movement disturbances in the workplace represents an undeniable human and financial challenge. Bureau of Labor Statistics (BLS, 2012) data show that, in the USA, these accidents represented about 30% of the 1,181,290 non-fatal occupational accidents (OA) with days lost in 2011. In its 2008 document on the "causes and circumstances of accidents at work in the European Union", the EC states that, among the 3,983,881 non-fatal accidents causing more than 3 lost work days in 2005, 19% were slips, trips, missteps, stumbles without a fall or with a fall on the level (CE, 2008). At companies operating under the French general social security system, slips, trips and other movement disturbances in work situations (excluding working at height) represented 32% of accidents with days lost (213,940 accidents); 34% of accidents with permanent partial disability (13,759 accidents); 35% of lost work days due to temporary disability (13,591,652 days) and 5% of fatal accidents (25 accidents) in 2011 (CNAMTS, 2012).

Analysis of this kind of accident is often limited to factors close to the injury in the accident genesis. However, accident analysis has also revealed explanatory factors distant from the injury, such as equipment usage (Kines, 2003), access system configuration (Leclercq et al., 2007), work system design (Derosier et al., 2008), work organization or safety management (Bentley and Haslam, 2001). Each factor revealed by analyzing an accident is required for its occurrence, irrespective of its position in the accident genesis. Investigating the accident genesis as far upstream of the injury as possible therefore assists prevention in terms of highlighting a maximum number and a variety of levers for action. A diversity of OAMD occurrence circumstances for different activity sectors (Leclercq and Tissot, 2004) has also been observed within a single company (Leclercq and Thouy, 2004). Combinations of factors common to several slips, collisions and other movement disturbances have been empirically identified in all the accidents subject to in-depth analysis at a regional power distribution facility (Leclercq and Thouy, 2004) and at a railroad company (Leclercq et al., 2007). The authors termed each of these combinations a "recurrent scenario".

* Corresponding author. Tel.: +33 383 509 866; fax: +33 383 502 185.
E-mail address: sylvie.leclercq@inrs.fr (S. Leclercq).

http://dx.doi.org/10.1016/j.aap.2014.04.004
0001-4575/

Haslam and Bentley (1999) had already observed that a combination of slippery conditions, use of footwear with worn treads and time-saving behavior was encountered in 50% of slip, trip and fall accidents among postal delivery workers.

Representing accidents by combinations of factors, rather than by isolated factors, allows us to characterize more closely accident-causing situations since an isolated accident factor (congested floor, person running, etc.) is more representative of a usual occupational situation than of an accident. Prevention is thus more a question of controlling factors, whose combination can be harmful, rather than trying to eliminate every risk factor (Monteau, 1997), which would indeed appear illusory in the case of OAMDs. Furthermore, a fact that has contributed to accident occurrence can sometimes only be considered an accident factor within the context of its occurrence. It may, in fact, be a safety-related factor in another context. For example, knowledge of a location is a safety-related factor, when a person anticipates a step (abrupt change in level) at a location where it is unusual (e.g. midway along a corridor). This same knowledge can be an unsafeness-related factor, when there is an unfamiliar obstruction and a person, trusting his/her knowledge of the location, does not notice it. Characterizing an accident by a combination of factors, such as an accident scenario, rather than by an isolated factor therefore allows us to consider contextual information reflecting the accident-causing nature of certain identified factors.

The purpose of this research is to develop serious recurrent OAMD scenarios, which go beyond the exclusively empirical stage of this development process adopted by Leclercq et al. (2007). Our work falls within the framework of a systemic accident model (Hollnagel, 2004), which has proved beyond any doubt its value to OAMD prevention (Bentley, 2009).

Bayesian Network (BN)-based approaches appear better suited to answering this kind of issue. They provide an adequate representation of our pre-processed data, being a set of accident factors combinations built by experts. Each combination is composed of qualitative knowledge (accident factors) and logical links (links between factors in each logical combination). BNs are well adapted to model such data, bridging the gap between different types of knowledge and unifying all available knowledge into a single type of representation. They are capable of apprehending qualitative knowledge, in terms of accident factors, and links through BN structure. They can also apprehend quantitative knowledge, in terms of frequency of accident factor occurrence among data set, through BN parameters, allowing recurrent scenarios extraction. Unlike other methods such as neural network models, regression methods etc., all the parameters in Bayesian networks have an understandable semantic interpretation. This method can therefore combine expert knowledge with data, to build the model. This is particularly useful when the amount of data is small. Moreover, if machine learning techniques are used (with or without expert knowledge) to build the model from a data set, it can be explained in terms that are understandable by domain experts.

BN-based occupational safety studies have been conducted by several authors in recent years. Using coded data, they have analyzed the effect of task performance-related factors in situations involving risks of falling from ladders or equipment such as scaffolding (Martín et al., 2009), the effect of safety climate- and individual experience-related factors on human behavior (Zhou et al., 2008) or the effect of accident factors (Zhao et al., 2012) or working conditions (García-Herrero et al., 2012) on accident occurrence. In the field of road accidents, BNs are increasingly used, e.g. to model and classify accidents according to their injury severity (Simoncic, 2004; Oña et al., 2011) or to predict the number of accidents of different severity (Deublein et al., 2013) or crash in real time (Hossain and Muromachi, 2012). To our knowledge, no BN-based research has investigated a methodology for determining recurrent scenarios as a diagnostic step toward improving occupational safety. This aim requires in-depth analysis of a set of accidents, which can be found in a database whose richest information is contained in narrative texts. Indeed, Lincoln et al. (2004) have shown that narrative text analysis is a useful supplement to traditional epidemiological analyses because it provides qualitative data, usually based on the accident/injury process, which offers a deeper understanding of the underlying accident process. Fatality investigation reports, in particular, contain data elements not routinely analyzed with coded occupational injury surveillance data (Bunn et al., 2008). The issue now is, "Is it possible to extract recurrent scenarios from a set of serious OAMD narrative texts?"

Further studies aimed at understanding accidents based on narrative text have been conducted in recent years. McKenzie et al. (2010) describe recent advances in using this kind of text in injury surveillance research. Narrative texts need to be pre-processed, unlike coded data which can be directly applied within the scope of BN-based approaches. Automatic methods, such as text mining, have been developed to extract clusters of words with a high probability of target category association (Brooks, 2008). However, these methods do not allow accurate identification of accident factors from a narrative text, i.e. facts that make sense in terms of the accident progression. Similarities or identities can effectively be expressed in words with different meanings or, conversely, a similar meaning can be expressed in differently spelt words (McKenzie et al., 2010). These facts can only be extracted, if the whole narrative is considered. To date, most OA analyses based on narrative text have implemented, for example a 'reconstruction template' (Lincoln et al., 2004), a priori-defined generic accident factors (Shibuya et al., 2010) or Haddon's matrix (Bunn et al., 2008) to process information manually. Analysis of these processed data is usually based on the occurrence and co-occurrence of factors or a number of related keywords.

Our aim is to extract combinations of factors common to several accidents, so we need to identify, from narrative texts, both accident generic factors, which have contributed to injury occurrence and how these factors have combined to cause injury. A BN-based approach has been developed to extract such combinations or recurrent scenarios.

2. Method

2.1. Data pre-processing

2.1.1. Data

The OAMD data used in this study were taken from the France's anonymous EPICEA database consolidating more than 18,000 OA cases that have occurred, since 1990, at companies operating within the French general social security system (EPICEA, 2011). EPICEA lists nearly all fatal occupational accidents and some accidents that were serious or significant for prevention. Identifying OAMDs contained in the database is not automatic. In particular, it requires analysis of the narrative text wording and reading of each account prior to its inclusion in the corpus data. Our study concentrated on the construction and metallurgical industries because these are the industrial sectors most affected by occupational accidents. These industries are dynamic and hazardous due to the diverse and complex nature of their work tasks, trades and environments, as well as the temporary and transitory nature of the workplaces and workforces (Kines, 2002). 143 accidents were ultimately extracted from EPICEA database, 79 cases from the construction sector and 64 cases from the metallurgical industries. However, this set is not representative of all OAMDs, so results could not be extrapolated to OAMDs occurring within the French general social security system. The construction and metallurgical industry set does allow us to develop the recurrent scenario extraction methodology and synthesize a set

of 143 OAMDs through a number of recurrent scenarios. Narrative texts were used because they offer a better understanding of the underlying accident process (cf. Section 1). The texts used were of variable length (30–337 words).

2.1.2. From narrative text to a logical combination of generic factors

Building recurrent scenarios requires us to extract, from a set of accidents, combinations of factors that have contributed to an injury and are common to several accidents. This means initially grouping singular accident factors, by definition specific to the accident with which they are associated, under the heading of generic factors; these then become factors common to the genesis of several accidents (Cuny et al., 2010). Accident-related dialog with the victim or investigator was impossible since accounts were extracted from a national anonymous database. Three OA experts identified descriptive passages expressing a fact that played a part in injury occurrence (singular factor) from the narrative text contained in each of the 143 accident reports. A fact was not retained as a singular accident factor, if the experts disagreed on its possible role in accident occurrence. Most of the singular factors retained were then consolidated into classes under the heading of generic factors. These generic factors were not predefined: they gradually emerged during reading and so they embrace, at best, all the accident factor-related information contained in the texts. Generic factor formulation and number stabilize, when a newly read accident record provides no further generic factor. Two generic factors, describing the victim's movement disturbance during the task and the injury-causing event, were systematically recorded from each text; these would be systematically part of the combination of generic factors describing the accident in question. Other selected generic factors were those embracing at least 5 singular accident factors, i.e. those for which the number of singular factors from the construction industry (Nconst) added to the one from the metallurgical industry (Nmet) is greater than or equal to 5.

A method inspired from the "causal tree" (Monteau, 1997) was applied to account for how the accident factors combined to cause injury. "Causal tree" method has been developed by the French National Research and Safety Institute (INRS – Institut National de Recherche et de Sécurité) to analyze singular occupational accidents. It is also called "INRS model" (Kjellén, 2000). It will be used here to combine the identified generic factors for each of the 143 considered accidents. This method starts from the injury causing event and involves asking "What event was necessary for its appearance?" and "Was another event necessary?" for the event considered and for each subsequently known event. This process was continued until all generic factors had been included in a logical combination ending in a succession of two events, namely the victim's movement disturbance during the task, which led to the injury causing event. Questioning enabled the contributing generic factors to be interlinked by a logical relationship.

Finally, each accident's narrative text was coded by a logical combination of generic factors. Fig. 1 illustrates an example of the OAMD coding step, in which each singular accident factor is highlighted in the narrative text and labeled according to its corresponding generic factor. A combination of several generic factors, whose number varies for each accident based on the relevant text content, is then identified for each accident using the method described above. To achieve our objective, a BN-based method was chosen to model OAMDs to ensure consideration of both accident factors and the links between them.

2.2. Bayesian network model for OAMD analysis

A BN is a directed acyclic graph (DAG) consisting of nodes and arrows, in which nodes represent random variables and arrows represent dependence relationships between connected nodes from a probabilistic standpoint. The novel representation for each accident includes generic factors as node values and their transitions as branches. The basic idea of such a logical combination is to consider each branch as a possible outcome of an event which is here a generic factor. Each node in a BN model has a specified conditional probability distribution (CPD); the model is parameterized by all the CPDs. One of the most important features of a BN model is factorization of joint probability space, so that conditional independence can be used to simplify modeling and save computations. A BN model is useful when combined with efficient algorithms for inference. Additionally, other inference tasks are performed, such as computing the K Most Probable Explanations (MPE) of evidence (K being an integer ≥ 1). Pearl (1988) and Pearl (2003) provide detailed information on BN. The steps followed in building a BN model for OAMDs are described below.

2.2.1. Nodes and node values

When building a BN model, the first task is to identify variables of interest (represented by nodes) and their possible states (represented by node values, i.e. here generic factors). All possible states must be mutually exclusive, or in other words each variable must take on exactly one of these states at a time. Common states in a discrete variable case are Boolean (True or False), ordered (low, medium, high) or integer (e.g. 1 to 10) values. In this study, variable states are sets of generic factors.

It should be noted that "mutually exclusive" means that the intersection of two sets is null, whichever 2 sets are considered among N sets. The universe is therefore partitioned into N sets, if these are mutually exclusive. To identify variables insuring this criterion, each generic factor was associated with one of four components representing the occupational situation and also with the level, at which this generic factor proved detrimental in the accident genesis. Indeed, accidents are phenomena whose origin and genesis lie in the operation of a specific system for producing goods or providing services, whose performance mechanism is recognized as complex (cf. e.g. Kjellén, 2000). So OAMD is here considered as a symptom of dysfunction of a system characterized by four interacting components (C1 to C4). Four levels in the accident genesis will be considered to distinguish disturbance in interacting components.

The four components identified from the set of generic factors are:

- C1, which represents the physical environment forming the injured worker's surroundings: machine, machinery, hand tools, apparatus, equipment, workplace, wind, snow, protection equipment, etc.;
- C2, which represents the injured worker's activity; this can be described with varying degrees of refinement. It can be movement or posture as well as displacement or manipulation of a tool;
- C3, which represents work organization; in this case materials provided, implemented means of collective protection/prevention, training or interference between activities;
- C4, which represents the injured worker; in this case experience in the workplace and unsteady operation due to a temporal context.

The four levels derived from the logical combinations of OAMD generic factors (see Fig. 2) are listed below from the injury to the most upstream level in the accident genesis:

- L1 represents the injury causing event.

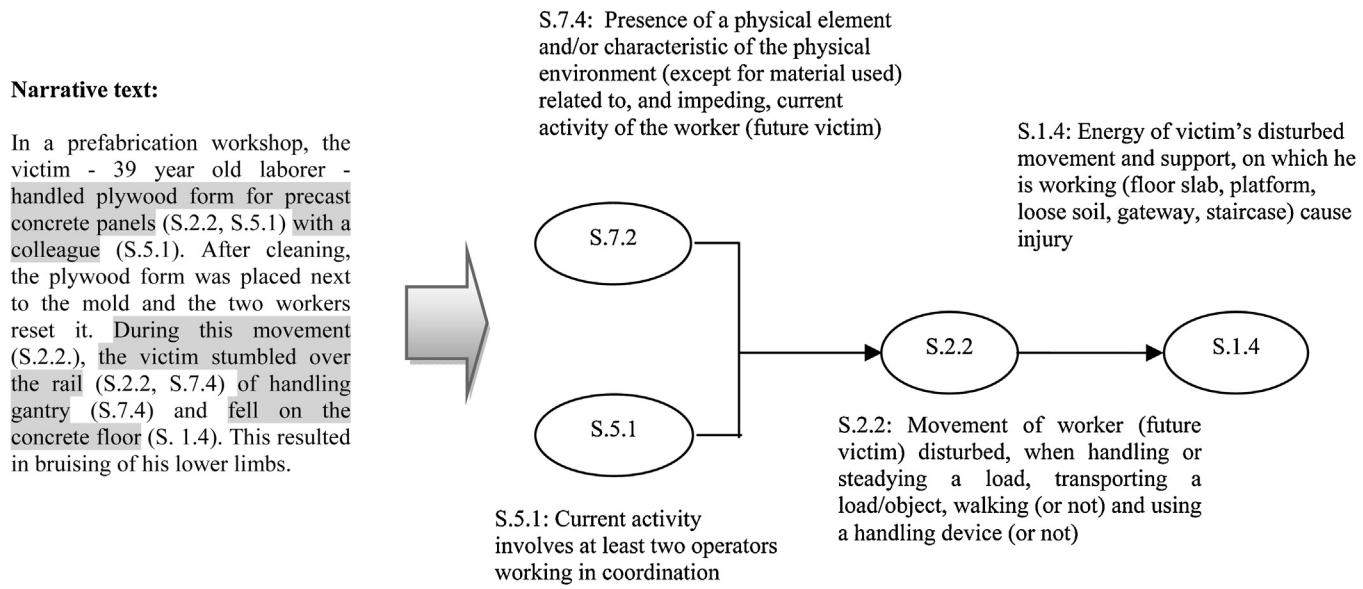

Fig. 1. An example of narrative text coding.

Fig. 2. OAMD genesis levels.

- L2 represents the victim's movement disturbance during the task, which can be conjoined with a factor expressing non-compliance with good working practice contributing to injury occurrence.
- L3 includes factors that have a direct impact on the movement performed in the accident causing situation. These factors provide a direct explanation of the movement disturbance.
- L4 includes factors that have an indirect impact on the movement performed in the accident causing situation. These factors provide an indirect explanation of the movement disturbance. Data is rarely assigned at this level.

Table 1 gives the 9 variables (*V1* to *V9*) and their possible states *S.x.y*, where *x* is the corresponding variable number and *y* is its state number. It shows for each state, the corresponding generic factor and the associated component and level. All variables are not observed for all accidents. Based on accident seriousness, the experts assumed that, if variables *V3*, *V8* and *V9* had contributed to injury occurrence, they should have been reported. So another state called "Nothing" will be added for the states of variables *V3*, *V8*, *V9*, which are henceforth considered as complete information.

2.2.2. BN model structure

There are two ways of building a BN structure using either structure learning algorithms or expert knowledge. The first method requires extensive data. To implement it in our case, the amount of data required would have been at least 155,520 accidents. This number was calculated using the expression: $3 * (\prod_{i=1}^{i=9} |V_i|)$, where $|V_i|$ is the cardinality of the ith variable. 150,000 observations were required instead of the 143 available. We therefore decided to use expert knowledge for building the BN structure, which was developed to embrace, at best, the 143 logical combinations of generic factors, including combinations formed by only three and those represented by more generic factors. A set of variables was proposed at each level of the structure illustrated in Fig. 3, which can be considered as a "maximum structure". In other words, several states for variables *V4* to *V7* were missing (no generic factor), when we encoded the 143 OAMDs based on this structure and, for the 143 OAMD cases, an accident is characterized by a number of generic factors that is less than the number of variables considered in the structure.

The variable at Level 1 (*V1*), which expresses the injury causing event, has a logical relationship with the variables at Level 2

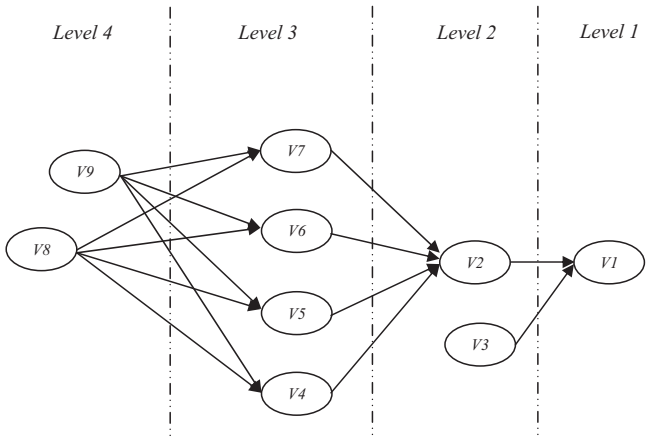

Fig. 3. OAMD Bayesian network structure.

Table 1
Variables (Var.) and their states. Each state corresponds to a generic factor, associated to a component (C1 to C4) and a level (L1 to L4), or to the state "Nothing". The right-hand column provides the number of singular factors consolidated under each referred generic factor from the construction (Nconst) and metallurgical (Nmet) industries (cf. Section 2.1.2).

Var.	Lev.	Comp.	Variable states (corresponding generic factors)	Nconst/Nmet
V1	L1	C1/C2	S.1.1: Moving heavy vehicle used in the workplace crushes victim	13/4
			S.1.2: Moving part of a machine, tool or device used in the workplace causes injury	15/18
			S.1.3: Element external to the victim (except a moving heavy vehicle and a moving part of a machine, tool or device) with which each contact results in injury: molten metal (zinc, aluminum), corrosive liquid (nitric acid), power component, tank of pasty wax	0/5
			S.1.4: Energy of victim's disturbed movement and support, on which he is working (floor slab, platform, loose soil, gateway, staircase) cause injury	22/4
			S.1.5: Energy of victim's disturbed movement and the item handled or worked by the victim or hand tool used by him cause injury	6/16
			S.1.6: Energy of victim's disturbed movement and an element of the physical environment with which each contact does not systematically result in injury (except support on which he works, the element handled or worked by him and the manual tool used by him) cause injury	23/17
V2	L2	C2	S.2.1: Movement of worker (future victim) disturbed, when performing a simple displacement	19/10
			S.2.2: Movement of worker (future victim) disturbed, when handling or steadying a load, transporting a load/object, walking (or not) and using a handling device (or not)	11/8
			S.2.3: Movement of worker (future victim) disturbed, when climbing up on, or down from, work equipment	16/3
			S. 2.4: Movement of worker (future victim) disturbed, when handling or manipulating an object/tool	27/42
			S.2.5: Movement of worker (future victim) disturbed, when waiting for or monitoring movement of a physical element or situation	6/1
V3	L2	C3	S.3.1: Activity performance without work equipment or when latter fails as required by regulations or good working practice	9/16
			S.3.2: Activity performance in absence of all protection/prevention conditions for the occupational situation (excluding work equipment) and required by regulations or good working practice (excluding work equipment-related conditions)	9/11
			S.3.3: Nothing	
V4	L3	C2	S.4.1: Worker (future victim) in an unstable posture	6/4
			S.4.2: Worker (future victim) makes unusual or unsuitable displacement, including the decision to move or not, how to move or displacement conditions and route	12/3
			S.4.3: Worker (future victim) uses material or equipment in unusual or unsuitable way with respect to regulations or good working practice	9/7
			S.4.4: Worker (future victim) performs an occasional activity not subsequent to an incident	8/10
			S.4.5: Worker (future victim) performs a recovery activity subsequent to an incident or to avoid being injured	8/10
			S.4.6: Worker (future victim) exerts an effort or forces against an element which resists	10/11
V5	L3	C3	S.5.1: Current activity involves at least two operators working in coordination	16/7
			S.5.2: Current activity involves at least two operators working in improvisation	2/5
			S.5.3: Training or qualification for the position lacking, insufficient or ongoing–except for safety training	6/2
V6	L3	C4	S.6.1: Worker (future victim) has less than three months experience in the workplace	13/8
			S.6.2: Worker (future victim) operating unsteadily because of temporal context (start or end of shift)	4/2
V7	L3	C1	S.7.1: Meteorological conditions make floor slippery	7/0
			S.7.2: Presence of a physical element and/or characteristic of the physical environment unrelated to, and impeding, current activity of worker (future victim)–except for cases in which meteorological conditions make floor slippery)	21/10
			S.7.3: Characteristic of material used makes activity of worker (future victim) difficult or impossible	19/13
			S.7.4: Presence of a physical element and/or characteristic of the physical environment (except for material used) related to, and impeding, current activity of the worker (future victim)	12/10
V8	L4	C1	S.8.1: Incident or accident involving physical environment disrupts current activity	8/16
			S.8.2: Nothing	
V9	L4	C3	S.9.1: Activity interferes in activity performed by worker (future victim) in a constraining way (or produces constraining elements)	6/2
			S.9.2: Nothing	

(*V2* and *V3*), which expresses the victim's movement disturbance during the task and non-compliance with good working practice contributing to injury respectively. *V2* has all the variables at Level 3 as parents. The victim's movement disturbance during the task can, for example, be caused by a variable related to the physical environment (*V7*), to any other variable at Level 3 (*V4* to *V6*) or to a conjunction of two, three or four of these variables (*V4* to *V7*).

The variables at Level 4 have all variables at Level 3 as children, which means that, if there is a technical incident factor (*S.8.1*) for example, this can cause unstable posture of the injured worker (*S.4.1*) or any other possible variable state or a conjunction of variable states at Level 3 (*V4* to *V7*).

Thus, each injury is the outcome of a set of variable states forming a specific configuration, based on the structure illustrated in Fig. 3.

2.2.3. BN inference

The structure of the BN corresponds to its qualitative aspect, while the inference in BN corresponds to the quantitative aspect which consists of computing probabilistic queries, taking into account all factors contributing to the occurrence of accidents. A junction tree inference engine (Jensen et al., 1990) was applied in this study in two steps. The first one consists in using graph theory to transform the initial graphical structure of the BN into a specific graphical entity called the junction tree. In the second step, the junction tree is used as a channel to transmit and propagate the effect of observations. For a BN, junction tree T consists of p cliques, say $C1,\ldots,Cp$ and $p-1$ separators, say $S1,\ldots,Sp-1$. The cliques correspond to a fully connected subset of nodes. The separators contain the information shared by a pair of adjacent cliques. The junction tree has a remarkable advantage, from the distributed computing

point of view, because in theory it can easily be split into different parts. The probability function f of the discrete variables x is defined as:

$$f(x) = \frac{\prod_{C \in \mathcal{C}} a_C(x_C)}{\prod_{S \in \mathcal{S}} b_S(x_S)} \quad (1)$$

where a_C and b_S are known non-negative real functions. Any non-negative function, which can be represented in the form of Eq. (1) will be said to factorize on T (Nilsson, 1998).

Jensen et al. (1990) propose a schedule consisting of two phases for the junction tree algorithm. They select an arbitrary clique *C1* as the 'root-clique'. An initial collection phase involving the passage of active flows only along edges toward *C1* is followed by a distribution phase in which, starting from *C1*, active flows are passed back toward the periphery. The message-passing protocol ensures that a clique can only send a message to a neighboring clique when it has received messages from all of its neighbors. After message passing, the potentials in the cliques are equal to the marginal probability of the nodes in the clique (given the evidence).

2.2.4. "Learning" the BN parameters

Bayesian model parameters (i.e. conditional probabilities) can be estimated, when the BN structure has been created and the states of variables can be obtained from the accident data. This is called parameter estimation or "learning" (Little and Rubin, 1997; Jensen and Nielsen, 2007). The Expectation–Maximization (EM) algorithm was used for this purpose. The EM algorithm (Lauritzen, 1995; Jensen and Nielsen, 2007) is a general approach for finding maximum-likelihood estimates for a set of parameters h, when researchers have an incomplete data set. The EM algorithm begins by randomly assigning a configuration h_0 to h_t when the algorithm is used in parameter learning, based on an outcome h_t after t iterations.

2.3. Scenario extraction approach

After building the BN model for the 143 serious OAMDs studied, the next crucial step was to extract recurrent scenarios, considering that certain variable states combinations were common to several accidents. OAMD clustering will group accidents based on these common patterns. Then, each of these patterns may be considered as the common part of a probable generic accident which is called recurrent scenario. The proposed approach to achieving this objective involves two steps, the first being a clustering process to gather similar accidents and the second being a scenario extraction from each cluster using the Most Probable Explanation (MPE).

As stated in Section 2.2.2, the OAMD model structure is considered as a "maximum structure" embracing all logical combinations of OAMD variable states. Several variable states are therefore missing in each accident for variables *V4* to *V7*. Data are considered along with these missing values for clustering purposes. However, when calculating the MPE for each cluster, the state "nothing" has been added to the states sets of variables *V4* to *V7*. In the BN model, this additional state, called "Nothing" or "False", means that this variable did not contribute to the occurrence of the OAMD in question. If the missing states had been considered in estimating the MPE, the extracted scenario would provide a state for each variable; i.e. the obtained scenario would contain 9 variable states that would not reflect the used data.

2.3.1. Clustering

The purpose of the clustering step is to group the OAMDs into "similarity" classes taking into account both variable states occurrence and the pattern they take for each accident. The number of clusters being unknown, the proposed clustering method is based on adding a node C to the structure illustrated in Fig. 3. The cardinality of this node corresponds to the number of clusters. Our aim being to extract patterns focusing on factors explaining movement disturbances, node C will be linked to the 4 variables at Level 3, i.e. variables explaining directly movement disturbance (cf. Fig. 2). Indeed, different injury events (*V1*) caused by different movement disturbances (*V2*) could be explained by similar generic factors combinations (*V4* to *V9* combined) (Leclercq et al., 2007).

A learning step with hidden variable (node C) should therefore be performed. The proposed model $M = (m, \theta)$ is a BN with variables $X_1, \ldots, X_n$; $n = 9$, if (1) structure m corresponds to a directed acyclic graph with nodes $X_1, \ldots, X_n$ and (2) parameters θ consist of probability distributions $P_M(X_i|pa(X_i))$, $i = 1, \ldots, n$, where $pa(X)$ is a set of parents of X in m. Then $P_M(X_1, \ldots, X_n) = \prod_i^n P_M(X_i|pa(X_i))$. The model gives a "soft" classification of d (OAMD). That is, d belongs to each class c_l with probability:

$$P(c_l|d) = \frac{P(c_l)P(d|c_l)}{P(d)} \quad (2)$$

where c_l specifies the state of the hidden variable (also called a latent class) (Oña et al., 2013).

For given data D, we therefore need to determine the optimal cardinality of the hidden variable and the model parameters. This is done by learning parameters for different cardinality of the hidden variable ($|C|$) and then selecting the $|C|$ that gives the best model based on some criterion (e.g. the Bayesian Information Criterion (BIC) score or the maximization likelihood). The optimal cardinality corresponds to the number of clusters providing the best separation between classes. In our study, we used the BIC (Raftery, 1986) to identify the model that adjusts best to the observations. In clustering contexts, the BIC performs better than other criteria (Biernacki and Govaert, 1999). Generally, the BIC is well suited to analyzing small samples because it is more sparing and adapts better to this kind of data. The lower the BIC score, the better the model. Depaire et al. (2008) and Oña et al. (2013) may be referred to for further information on clustering using a BN.

2.3.2. Most Probable Explanation

After the clustering step, each accident is associated to one of the $|C|$ clusters. For each cluster, a structure derived from the one on Fig. 3 has been considered to estimate the parameters as accurately as possible. Variables cardinalities have been reduced to take into account only the states appearing in the cluster for each variable.

A recurrent scenario is considered the most probable configuration of variable states for each cluster. This can be thought of as providing a plausible "explanation" for the observations in a cluster and is called a Most Probable Explanation or MPE (Nilsson, 1998; Cowell et al., 1999). A MPE in a BN is the complete variable instantiation with the highest probability given current evidence (Pearl, 1988).

An important point concerning the chosen algorithm (MPE) is that it does not intuitively accept the most likely state for a given variable, but makes a decision based on the entire sequence of variable states, favoring here the links between them. Thus, a particularly "unlikely" variable state midway through the sequence will not matter, provided the overall context of the data observed is reasonable. These aspects connected to MPE algorithm have been analyzed in depth in other research fields, such as medical diagnosis (Lucas et al., 2000) or economic processes (Demirer et al., 2006).

The junction tree algorithm becomes an MPE algorithm by replacing the sum (sum-marginalization) by the maximum (max-marginalization) in the propagation algorithm. The junction tree guarantees the existence of a configuration v. Furthermore,

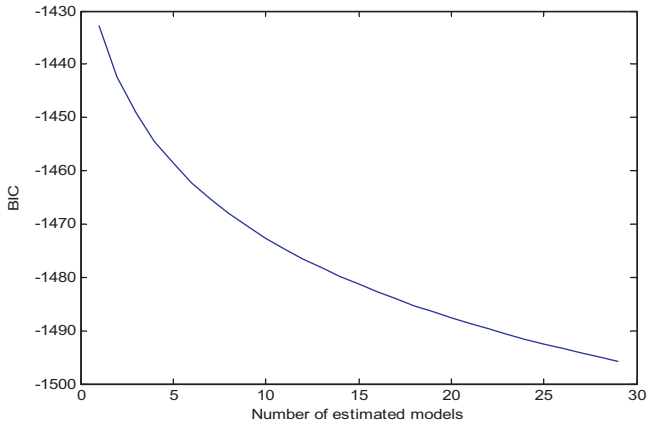

Fig. 4. BIC score according to the class cardinality.

ν maximizes f since the max-marginal charge is a representation for f (Nilsson, 1998), such that

$$f(\nu) = \frac{\prod_{C \in \mathcal{C}} \hat{f}_C(x_C)}{\prod_{S \in \mathcal{S}} \hat{f}_S(x_S)} = \max f \tag{3}$$

This complete observation ν is a scenario representative of the most probable variable states combination among a set of similar accidents. It is considered the recurrent scenario in this case.

3. Results and discussion

3.1. Clustering step

Fig. 4 shows the fall in the BIC score based on the number of clusters considered from the 143 OAMD accounts. Increasing the number of clusters reduces the BIC values, but a higher number of clusters implies a higher degree of complexity. From a practical point of view, a marginal improvement in statistical fit is not that useful, when a much higher degree of complexity is introduced.

In the literature, Depaire et al. (2008) selected the model, in which the BIC and the CAIC showed virtually no further improvement. Scheier et al. (2008) chose a model, in which the differences between two successive values of the BIC or the CAIC were less than 1%. Taking inspiration from the literature, 8 clusters were chosen in our research as a compromise between statistical fit and clustering structure complexity. Indeed the difference between BIC values for the 8th and the 9th estimated models is less than 1%.

Fig. 5 depicts the OAMD BN proposed for the clustering step. It is composed of the 9 variables shown in Fig. 3 and the node *C* defined by 8 states. It becomes a valuable estimating tool, once the network has been structured and the distribution parameters have been learnt. Each accident characterized by the 9 variables is considered an observation, for which the marginal probability of the node *C* is calculated.

The percentage of OAMDs, which occurred in each activity sector, is shown in Table 2 for each of the 8 clusters. These percentages are variable. Two clusters out of eight are even from one sector only. This means that the generic factor combinations are more or less frequent in relation to serious OAMDs in each sector. If we examine the distribution of generic factors among the OAMDs studied (cf. Table 1), we observe that most of these factors are shared by both sectors. However the number of singular factors from each sector (Nconst and Nmet from construction and metallurgical sectors respectively) consolidated under a generic factor was more or less important depending on the generic factor. Two factors are specific to one sector, namely S.7.1 and S.1.3. The level of generality applied (e.g., the same "claw hammer" object can be expressed by "hammer", "hand tool" or "tool" with increasing levels of generality) sometimes prevented emergence of a generic factor specific to an occupational sector. The level of generality is especially determined by the minimum number of singular factors (5) consolidated under a generic factor, from which this generic factor is selected. This sometimes leads to consolidating several generic factors under a higher level of generality heading, which lead to a combination of a number of generic factors, one of which may be specific to a sector.

3.2. Scenario extraction step

For each cluster, a learning step was repeated to update the parameters of the associated model and to derive the MPE. Fig. 6 illustrates MPE results for cluster 2. The variable states retained are displayed as black bars. We note that the black bars do not correspond systematically to the probability maximum. The aim of MPE method is to favor the observation that maximizes the flow of information in the network, but not each variable separately. State *S.2.1* is the third that involves occurrence of variable *V2*. This was observed more frequently with the combination *S.4.2*, *S.7.2* and *S.9.1* for variables *V4*, *V7* and *V9* respectively. Moreover, we note that all transitions between variables have been observed, except for the transition between *V2* and *V1*. Among the 27 accidents in Cluster 2, the injury causing fact in this scenario (S.1.2) never follows the movement disturbance (S.2.1) in the scenario (see Table 2 and Fig. 7). The same phenomenon was noted with scenarios 6 and 7 for variables *V1* and *V2*. Observing all clusters reveals that clusters 2, 6 and 7 feature two different configurations emerging from each cluster. This highlights the non-uniformity of the OAMDs in these clusters, which has affected MPE calculation. The most probable explanation (Nilsson, 1998) would therefore be explored in future work.

The state "nothing" has been added for the variables *V4* to *V7* when evaluating MPE. This state is very frequent in the observed data. So accident scenarios can reveal it even if another state (a generic factor) is observed several times in the corresponding cluster. In this sense, the extracted scenarios reflect the data used, when the observed accident factors combinations involve less than 9 factors. We in fact calculated the MPE assuming complete data to obtain an estimate, which reflects the OAMD observations used (cf. Section 2.3).

Table 2 consolidates results for the 8 scenarios, each of which is specifically characterized by both its title and its variable states structure. Only the outcome states corresponding to generic factors, i.e. without showing the state "Nothing" are displayed. Each state is placed in the tree at the position of its variable in the structure (see Fig. 3).

Scenario 2, illustrated in Fig. 7, is the only one containing a conjunction of two generic factors, although conjunctions of at least two generic factors occur in 42 of the 143 OAMDs just before movement disturbance. These conjunctions clearly provide a more accurate explanation of the movement disturbance than chains of factors. The variety of injury-causing situations in the data makes it difficult to prompt emergence of similar conjunctions.

This scenario highlights the fact that obstruction to movement in an occupational situation is not always a permanent factor but can be the result of an activity different to the activity performed by the victim. In many other cases of OAMDs, physical elements or characteristics of the physical environment impeding the victim's current activity are related to this activity (cf. Table 1). In both cases, it is clear that prevention cannot simply provide a general recommendation to "remove obstructions to displacement".

The unusual or unsuitable displacement referred to in this scenario raises questions on the reasons underlying it and this effectively highlights the lack of useful information for understanding accidents in many narrative texts. Going back to the 143 texts,

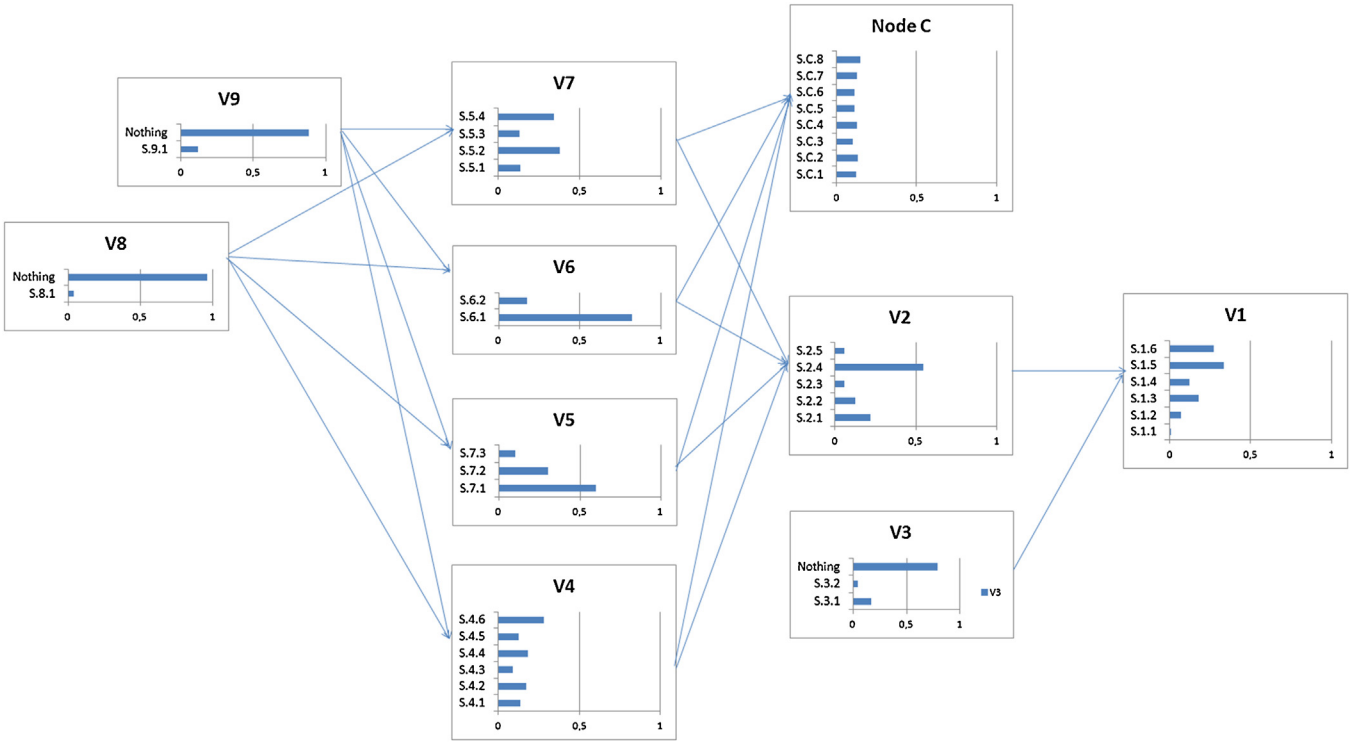

Fig. 5. Estimated occurrence probability of each variable states based on survey data.

we can provide an explanation in some cases: an incident, a characteristic of the material used, information given to the worker or worker experience. This scenario corresponds to any of the 27 OAMDs in cluster 2: we find factor combination *S.9.1*/*S.7.2*/S.4.2 and *S.2.1* in one case, combination *S.9.1*/*S.7.2*/S.2.1 in three cases and the state "nothing" in all cases except one for *V6* and *V8*.

Scenario 5, shown in Fig. 8, is one of the deeper scenarios, in which an incident disrupts an activity and leads to the injury. The scenario occurs only in metallurgy (11% of OAMDs from this sector) and corresponds to 2 of the 7 OAMDs in Cluster 5, in which we find combination *S.8.1*/*S.4.5* and *S.2.4* in 3 cases and the state "nothing" in all cases except one for *V9*.

Table 2
Summary of results for extracted scenarios. (%const-%met) is % of OAMDs from construction and metallurgy sectors respectively, contained in each cluster. Arrows have been crossed when the corresponding transition has not been observed in the cluster concerned.

Scenario number	%Const-%Met	Scenario heading	Variable states combination
1	14% – 19%	Exerting an effort against an element which resists	S.4.6 → S.2.4 → S.1.4 S.9.1 → S.7.2 → S.2.1 ✗ S.1.2 S.4.2 → S.2.1
2	25% – 11%	An element which is a product of activities other than that performed by the victim impedes his activity	
3	4% – 0%	When climbing up on, or down from, work equipment	S.7.2 → S.2.3 → S.1.4
4	15% – 11%	A moving heavy vehicle crushes the victim on the ground because his movement has been disturbed while handling an object with a colleague	S.5.1 → S.2.2 → S.1.1
5	0% – 11%	Following a technical incident, movement of an operator manipulating a tool is disturbed, the tool thereby causing injury	S.8.1 → S.4.5 → S.2.4 → S.1.2
6	8% – 1%	Following a technical incident, movement of an operator who is climbing up on, or down from, work equipment is disturbed	S.8.1 → S.4.5 → S.2.3 ✗ S.1.2
7	15% – 12%	When walking, an operator stumbles against an obstacle not involved in his activity	S.7.2 → S.2.1 ✗ S.1.5
8	19% – 34%	An operator is burned by a corrosive liquid after his movement was disturbed when handling an object	S.2.4 → S.1.3

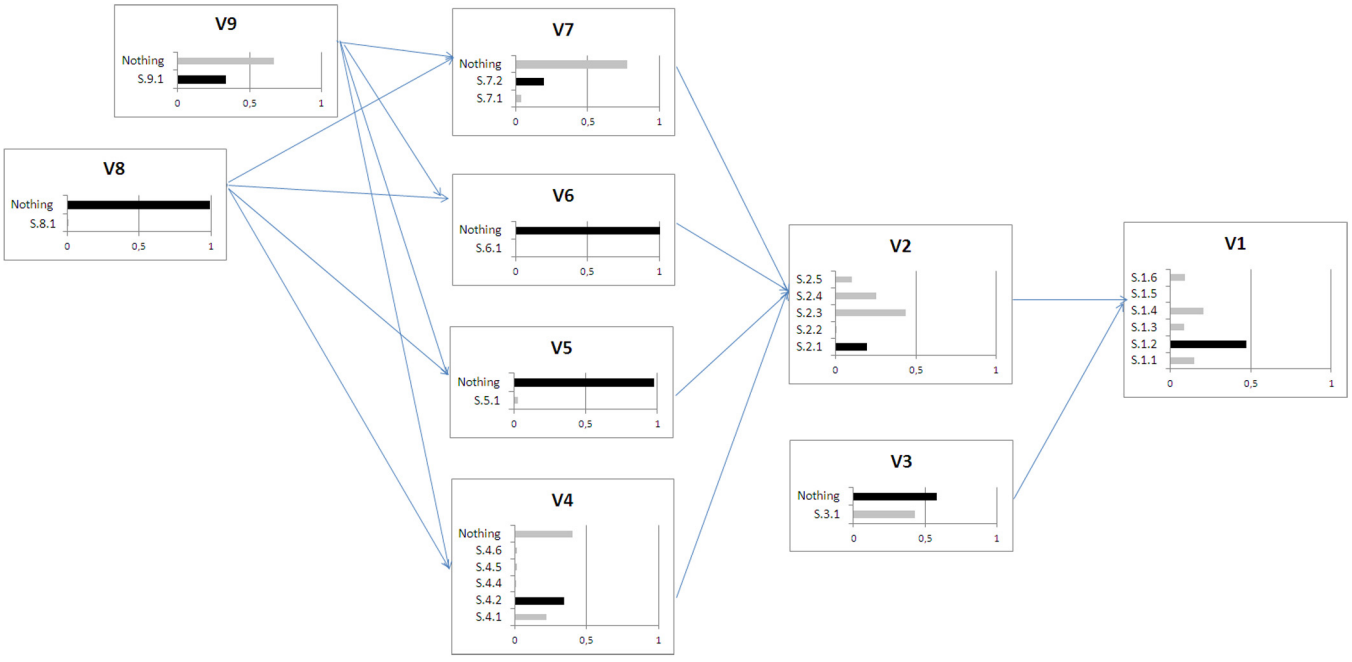

Fig. 6. Example of retained factors (black bars) for cluster 2 after parameter learning.

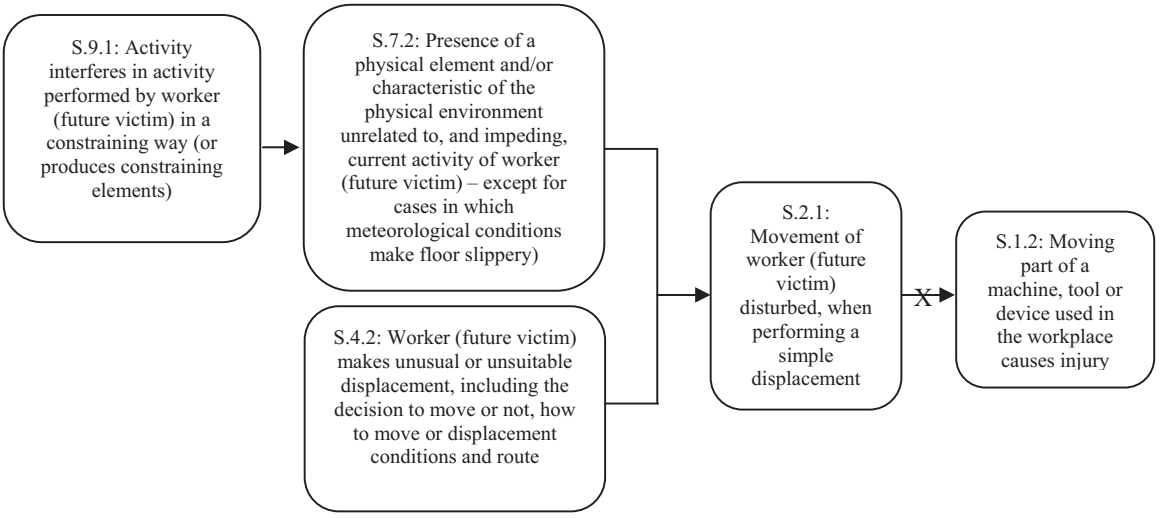

Fig. 7. Scenario 2 – An element, which is a product of activities other than that performed by the victim, impedes his activity.

Scenario 7, illustrated in Fig. 9, recalls the most common representation of accidents with movement disturbance: a person who walks slips or trips and then falls. However, it was extracted from only 20 OAMDs out of the 143. Amongst the 20 accidents in the cluster, the injury causing fact in this scenario never follows the movement disturbance generic factor in the scenario. The most frequent injury causing factor in Cluster 7 is S.1.6 (cf. Table 1), which is more consistent with the common representation of this type of accident. This important aspect of the results, involving to the methodology, has been discussed earlier. The scenario corresponds to any of the 20 OAMDs in the cluster, in which we find combination S.7.2, and S.2.1 in 7 cases and the state "nothing" in all cases except one for *V9*.

Scenario 8, illustrated in Fig. 10, is the shortest scenario in terms of the number of generic factors and can be explained by the highest number of accidents in Cluster 8, i.e. greatest variety of

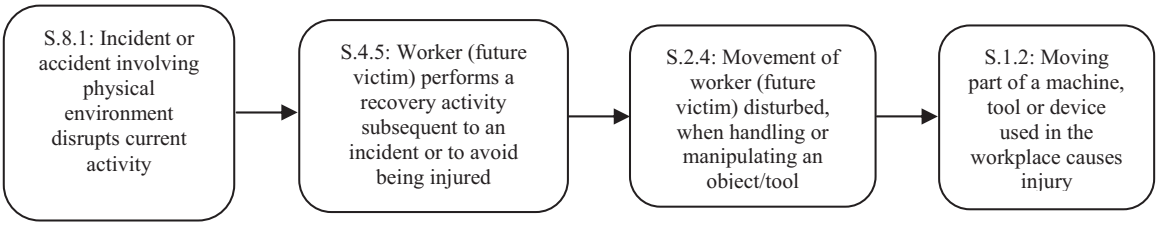

Fig. 8. Scenario 5 – Following a technical incident, movement of an operator manipulating a tool is disturbed, the tool thereby causing injury.

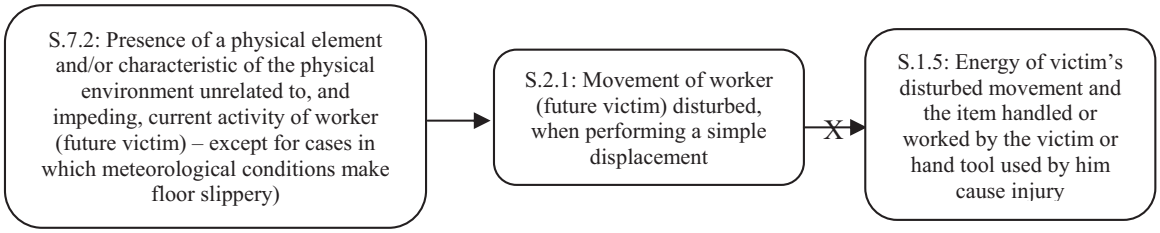

Fig. 9. Scenario 7 – When walking, an operator stumbles against an obstacle not involved in his activity.

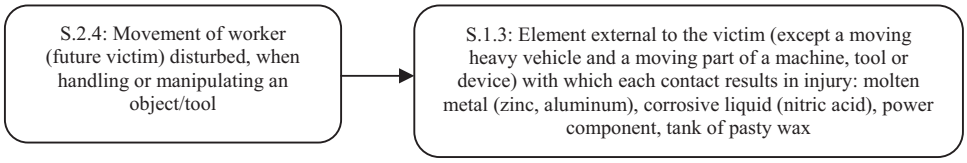

Fig. 10. Scenario 8 – An operator is burned by a corrosive liquid after his movement was disturbed when handling an object.

injury-causing situations, which make it difficult to prompt emergence of multiple similar factors. Consequently, the state "Nothing" appears here to be the most frequent one in the scenario. 42 OAMDs (amongst 64) and 27 (amongst 79) occurred when handling or manipulating an object or a tool in the metallurgy and construction sectors respectively. This scenario corresponds to 1 of the 37 OAMDs in the cluster, in which we find the state "nothing" in all cases for *V8* and *V9*.

The results of our study provide a description of the variety of circumstances, in which serious OAMDs occur. The eight extracted scenarios reflect accident narrative texts in the construction and metallurgy sectors. Two out of the 30 generic factors are specific to one sector and two of the eight extracted scenarios are specific to one sector only. The diversity of activities and companies in the data used and the low number of narrative texts in relation to this diversity make it more difficult to extract recurrent scenarios. Derived scenarios are most probable scenarios among accidents consolidated into the same class, so they could be considered similar.

Scenarios 2, 5 and 6 highlight the impact of technical incidents, accidents and coactivity respectively in the OAMD occurrence, a connection that is rarely made in the literature for OAMDs.

The derived scenarios differ from those referred to in the introduction, which were empirically extracted (Leclercq and Thouy, 2004; Leclercq et al., 2007). In these earlier studies, extracted accident factor combinations were common to a set of accidents considered accordingly as a cluster. In the present study, the clustering step can only precede the accident factor combination extraction. It is noted that each combination is partly shared by most, or several, accidents in the cluster. Factors additional to this shared combination are more or less frequent in the cluster. Several reasons can explain the differences between empirical scenarios and the results of the current work:

- In this study, the accidents analyzed occurred at different companies in two sectors, whereas in the previous study, accidents occurred at only one multitrade company.
- In this study, information explaining the accidents was extracted from narrative texts, whereas in the previous study, information was extracted from accident analysis and task analysis.
- In this study, all accidents analyzed were clustered into classes, whereas in the previous study, some accidents were consolidated based on their similarities, while others were left out.

Scenarios derived in this study would therefore be more consistent with the definition provided by Khan and Abbasi (2002, p. 468) that "A scenario is neither a specific situation, nor a specific event, but a description of a typical situation that covers a set of possible events or situations" than with the recurrent scenarios extracted empirically by Leclercq et al. (2007). However, if the extracted scenarios are not recurrent scenarios, the authors formulate the hypothesis that this puts into question more the data uniformity than the MPE method used. Finally, the uniformity and richness of the accident analysis results would enable us to enhance the accuracy, richness and recurrence of scenarios. While knowledge of different types of injury-causing accident factor is helpful for prevention, knowledge of recurrent combinations of factors is itself required for identifying OAMD-causing situations. In these accident cases, the related risk does not always involve an element with which contact invariably causes injury, such as high voltage. This knowledge helps in identifying actions which should be performed in priority in relation to preventing frequent accidents involving all workers, without exception.

4. Conclusions and future research

This paper describes a probabilistic approach to extracting recurrent scenarios from serious Occupational Accidents with Movement Disturbance (OAMDs). Analysis of this type of accident is often limited to factors close to the injury in the accident genesis. However, the causality of such accidents, indeed of any occupational accident, originates in a specific production activity context. A fundamental notion in this study is the representation of accidents by combinations of factors, rather than by isolated factors in order to characterize more closely accident-causing situations. Such combinations, common to several accidents, have been empirically identified at companies and are considered to be recurrent scenarios. In relation to preventing frequent accidents involving all workers, without exception, the mere fact that several factors contribute to the occurrence of the vast majority of such accidents is insufficient for characterizing risk situations. Their reconstitution, in the form of recurrent scenarios, is essential to more effective prevention.

The proposed approach focuses on identifying factor combinations common to several accidents extracted from narrative text. It comprises four parts: a necessary initial coding step to extract relevant information from text data. Each accident narrative text is

coded by logically combining generic factors inspired by the INRS model (Monteau, 1997). It should be noted that this coding step is very time consuming for experts and could not be performed on a large data set. A Bayesian Network (BN)-based model is then built for OAMDs using expert knowledge and data extracted from 143 narrative texts, which combine qualitative and quantitative aspects of the relevant knowledge. Following these initial steps, the key step in our work involves clustering OAMDs into "similarity" classes taking into account both the generic factors occurrence and pattern. Finally, the Most Probable Explanation (MPE) is derived for each cluster.

Expert knowledge is therefore the essential foundation of the two initial steps in order to make up the small sample used. Consequently, validity of the identified scenarios could be confirmed by adopting BN structure learning process. This would require the use of many more cases of serious OAMDs. Moreover, these cases should be characterized not only by factors describing the injuries but also by factors explaining movement disturbances. In a context where these accidents are rarely analyzed in depth, this is a real difficulty. We successfully used a BN to represent OAMDs and extract OAMDs scenarios. Scenario richness depends on the depth of analysis, the uniformity of circumstances, in which accidents occur, and the level of generality of the generic factors. The results of our study are useful in the prevention field on the one hand for generating company awareness of the variety of circumstances, in which these accidents occur, and, on the other hand, for developing an analysis framework dedicated to this type of accident. The analysis levels defined here could be useful for developing such a framework which could also be used to get homogeneous data on OAMDs and so more precise scenarios.

The present work allowed us to answer the issue stated in the introduction "Is it possible to extract recurrent scenarios from a set of serious OAMD narrative texts"? Two occupational sectors have been considered here, based on the hypothesis according to which generic factors and genesis of this kind of accidents may be different in different activity sectors. The results support this hypothesis. Nevertheless, they may be even less generalizable to all OAMDs that only particularly serious OAMDs occurred in two sectors have been analyzed.

In the future, logical combinations of generic factors derived from accidents analyzed more deeply and containing less missing values, may enable the dynamic BN to be used to obtain a more accurate, comprehensive representation of OAMD genesis based on the model's dynamic aspect. This characteristic allows us to repeat the same structure in different slices, when OAMD representation may be considered in levels and slices. In other words, these slices could enable us to consider sequences of factors (or of conjunctions of factors) instead of just one factor (or conjunction of factors) at the same level of the OAMD model. Moreover, other improvement could be considered in the process such as the cross-validation for the assessment of the model accuracy.

Lastly, this study takes into account the factors, considered by three experts as having had a role in accidents occurrence. What about the strength of the link between them? Causal relationships are indisputable between some of them, especially technical factors. For example, there is no doubt a slippery floor caused a slipping when accident analysis revealed a link between these two facts. However a slippery floor is not enough to lead to a slipping. Individual factors such as tiredness, experience or organizational factors leading to precipitation for example can combine with a slippery floor to cause the slipping. Even if most often technical component cannot be the only cause of movement disturbance, it is not possible to offer evidence for causal relationships among factors when individual or organizational factors are involved. The idea here is to search for recurrent combinations of factors of different nature. More often the involvement of individual and organizational factors in such combinations, more strong the "causal" link between them and other factors.

Acknowledgements

Authors would like to thank Professor Philippe Leray at the University of Nantes.

References

Bentley, T., 2009. The role of latent and active failures in workplace slips, trips and falls: an information processing approach. Appl. Ergon. 40, 175–180.
Bentley, T., Haslam, R., 2001. A comparison of safety practices used by managers of high and low accident rate postal delivery offices. Safety Sci. 37, 19–37.
Biernacki, C., Govaert, G., 1999. Choosing models in model-based clustering and discriminant analysis. J. Stat. Comput. Sim. 64, 49–71, http://dx.doi.org/10.1080/00949659908811966.
BLS, 2012. [Online] Available at: http://www.bls.gov/iif/#tables (accessed 15.01.13).
Brooks, B., 2008. Shifting the focus of strategic occupational injury prevention: mining free-text, workers compensation claims data. Safety Sci. 46 (1), 1–21.
Bunn, T.L., Slavova, S., Hall, L., 2008. Narrative text analysis of kentucky tractor fatality reports. Accid. Anal. Prev. 40 (2), 419–425.
CE, 2008. [Online].Available at: http://*ec.europa.eu/social/BlobServlet?docId=2785&langId=fr*.[Accessed 19 September 2013].
CNAMTS, 2012. Statistiques Nationales des accidents du travail, des accidents de trajet et des maladies professionnelles. CNAMTS, Paris.
Cowell, R.G., Dawid, A., Lauritzen, S., Spiegelhalter, D.J., 1999. Probabilistic Networks and Expert Systems. Springer-Verlag, New York.
Cuny, X., Monteau, M., Leclercq, S., 2010. The typical scenario: towards extension of STF analysis. In: International Conference on Fall Prevention and Protection-NIOSH, Morgantown, USA.
Demirer, R., Mau, R., Shenoy, C., 2006. Bayesian Networks: a decision tool to improve portfolio risk analysis. J. Appl. Financ. 16, 106–119.
Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accid. Anal. Prev. 40 (4), 1257–1266.
Derosier, C., Leclercq, S., Rabardel, P., Langa, P., 2008. Studying work practices: a key factor in understanding accident on the level. Ergonomics 51 (12), 1926–1943.
Deublein, M., Schubert, M., Adey, B.T., Köhler, J., Faber, M.H., 2013. Prediction of road accidents: a Bayesian hierarchical approach. Accid. Anal. Prev. 51, 274–291.
EPICEA, 2011. [Online] Available at: http://www.inrs.fr/accueil/produits/bdd/epicea.html (accessed 01.10.11).
García-Herrero, S., Mariscal, M., García-Rodríguez, J., Ritzel, D.O., 2012. Working conditions, psychological/physical symptoms and occupational accidents. Bayesian network models. Safety Sci. 50 (9), 1760–1774.
Haslam, R.A., Bentley, T.A., 1999. Follow up investigations of slip, trip and fall accidents among postal delivery workers. Safety Sci. 32, 33–47.
Hollnagel, E., 2004. Barriers and Accident Prevention. Ashgate Publishing Limited, Hampshire (Royaume Uni).
Hossain, M., Muromachi, Y., 2012. A Bayesian network based framework for real-time crash prediction on the basis freeway segments of urban expressways. Accid. Anal. Prev. 45, 373–381.
Jensen, F.V., Nielsen, T.D., 2007. Bayesian Networks and Decision Graphs, 2nd ed. Springer-Verlag, New York:.
Jensen, F.V., Olesen, K.G., Andersen, S.K., 1990. An algebra of Bayesian belief universes for knowledge-based systems. Networks 20 (5), 637–659, http://dx.doi.org/10.1002/net.3230200509.
Khan, F.I., Abbasi, S.A., 2002. A criterion for developing credible accident scenarios for risk assessment. J. Loss Prevent. Proc. Ind. 15 (6), 467–475.
Kines, P., 2002. Construction workers' falls through roofs: fatal versus serious injuries. J. Safety Res. 33, 195–208.
Kines, P., 2003. Case studies of occupational falls from heights: cognition and behavior context. J. Safety Res. 34, 263–271.
Kjellén, U., 2000. Prevention of Accidents Through Experience Feedback. Taylor and Francis, London.
Lauritzen, S., 1995. The EM algorithm for graphical association models with missing data. Comput. Stat. Data Anal. 19 (2), 191–201, http://dx.doi.org/10.1016/0167-9473(93)E0056-A.
Leclercq, S., Thouy, S., 2004. Systemic analysis of so-called accident on the level in a multi trade company. Ergonomics 47 (12), 1282–1300.
Leclercq, S., Tissot, C., 2004. Serious falls on the level in occupational situations. In: Mc Cabe, P.T. (Ed.), Contemporary Ergonomics. CRC Press, London.
Leclercq, S., Thouy, S., Rossignol, E., 2007. Progress in understanding processes underlying occupational accidents on the level based on case studies. Ergonomics 1 (15), 59–79.
Lincoln, A.E., Sorock, G.S., Courtney, T.K., Wellman, H.M., Smith, G.S., Amoroso, P.J., 2004. Using narrative text and coded data to develop hazard scenarios for occupational injury interventions. Inj. Prev. 10, 249–254, http://dx.doi.org/10.1136/ip.2004.005181.
Little, R., Rubin, D., 1997. Statistical Analysis with Missing Data. Wiley & Sons, John New York.
Lucas, P., De Bruijn, N., Schurink, K., Hoepelman, A., 2000. A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. Artif. Intell. Med. 3, 251–279.

Martín, J.E., Rivas, T., Matías, J.M., Taboada, J., Arguelles, A., 2009. A Bayesian network analysis of workplace accidents caused by falls from a height. Safety Sci. 47 (2), 206–214.
McKenzie, K., Scott, D.A., Campbell, M.A., McClure, R.J., 2010. The use of narrative text for injury surveillance research: A systematic review. Accid. Anal. Prev. 42 (2), 354–363.
Monteau, M., 1997. Analysis and reporting accident investigation. In: Encyclopedia of Occupational Health and Safety 1. BIT, Genève, pp. 57.22–57.25.
Nilsson, D., 1998. An efficient algorithm for finding the M most probable configurations in probabilistic expert systems. Stat. Comput. 8, 159–173.
Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian network. Accid. Anal. Prev. 43, 402–411.
Oña, J., López, G., Mujalli, R., Calvo, F.J., 2013. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. Accid. Anal. Prev. 51, 1–10.
Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Mateo, CA.
Pearl, J., 2003. Statistics and causal inference: a review. Test J. 12 (2), 101–165.
Raftery, A., 1986. A note on Bayes factors for log-linear contingency table models with vague prior information. J. Roy. Stat. Soc. 48 (2), 249–250.
Scheier, L.M., Abdallah, A.B., Iniardi, J.A., Copeland, J., Cottler, L.B., 2008. Tri_city study of Ecstasy use problems: a latent class analysis. Drug Alcohol Depend. 98, 249–263.
Shibuya, H., Cleal, B., Kines, P., 2010. Hazard scenarios of truck drivers' occupational accidents on and around trucks during loading and unloading. Accid. Anal. Prev. 42 (1), 19–29.
Simoncic, M., 2004. A Bayesian network model of two-car accidents. In: Bauer, L. (Ed.), Proceedings of the 22nd International Conference on Mathematical Methods in Economics. Czech Republic, pp. 282–287.
Zhao, L., Wang, X., Qian, Y., 2012. Analysis of factors that influence hazardous material transportation accidents based on Bayesian networks: a case study in China. Safety Sci. 50, 1049–1055.
Zhou, Q., Fang, D., Wang, X., 2008. A method to identify strategies for the improvement of human safety behavior by considering safety climate and personal experience. Safety Sci. 46 (10), 1406–1419.