A CASE-BASED REASONING APPROACH TO CONSTRUCTION SAFETY RISK ASSESSMENT

GOH YANG MIANG

NATIONAL UNIVERSITY OF SINGAPORE

2004

A CASE-BASED REASONING APPROACH TO CONSTRUCTION SAFETY RISK ASSESSMENT

GOH YANG MIANG

(*B.Eng.(Hons.), NUS*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CIVIL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2004

ACKNOWLEDGEMENT

This research would not have been possible if not for the help of numerous people. Most importantly is the support and understanding that my wife and my family gave me throughout the PhD candidature. Without their support (especially my wife's) I would not have sustained through the stress and frustration of this research.

I would also like to express my appreciation for the guidance given by my academic supervisor, Associate Professor David Chua. I thank him for the frequent discussions and valuable advice. I am also grateful that despite his busy schedule he managed to allocate time to read and comment critically on my thesis.

This thesis also utilised a large amounts of input from industry experts and practitioners. These people include, Mr. Lim Poo Yam and his colleagues of Land Transport Authority (LTA), Mr. Ho Siong Hin and his colleagues of the Ministry of Manpower (MOM), Mr. Harry Ho, Mr. Chin Chee Chow and their colleagues of Singapore Construction Safety and Consultancy (SC2) Pte Ltd, Mr. Tan Kai Hong and his colleagues of SembCorp Engineers and Constructors, and Mr. Jason Oh of IES, EEHS Technical Committee. Special acknowledgement is also given to UK safety expert, Mr. John Anderson, who advised me on various issues relating to the research. I am also very grateful to all the individuals that helped the research in one way or another, but they were not named due to space constraints or their request for anonymity.

I also hope that the ideas proposed in this thesis will bring about improvement to the level of safety on construction sites and help prevent unnecessary loss of lives.

I dedicate this thesis to my baby girl, Goh Yu Le, who has brought much joy into my life, and my wife, Khew Hui Fong, for her unwavering support.

TABLE OF CONTENTS

SUMMARY

The construction industry is renowned for its poor safety records. One of the main strategies that can help to improve the safety performance of the industry is to ensure continual improvement of project safety management systems (SMS). This research proposes two levels of safety knowledge feedback that can facilitate the continual improvement of SMS. The first level of feedback refers to effective and thorough incident investigation after incident occurrence. The incident investigation should lead to an evaluation and improvement of the SMS that had failed and caused the incident. The second level of feedback is focused on ensuring that valuable safety knowledge in the form of safety plans and incident investigation reports are made available and useable for new project safety planning processes. Effective implementation of the second level of feedback would facilitate transfer of safety knowledge across projects and learning from past mistakes.

To facilitate the two levels of feedback, this research developed an incident causation model, known as the Modified Loss Causation Model (MLCM), which can be used to structure a thorough incident investigation process (first level of feedback) and act as a knowledge framework that facilitates the feedback of safety knowledge during new project safety planning (second level of feedback). The MLCM had been developed based on an in-depth literature review and evaluation of 140 actual accident cases obtained from Singapore's Ministry of Manpower.

To realize the second level of feedback, a novel case-based reasoning (CBR) approach of risk assessment was developed. The CBR approach was designed to facilitate the Job Hazard Analysis (JHA) method of risk assessment so that the approach is aligned

with the norm of structuring construction project plans based on activities. The key components of the CBR approach are: (1) a detailed MLCM-based knowledge representation scheme that can be used to capture and abstract key safety knowledge from incident cases and past risk assessments, (2) a case retrieval mechanism based on customized similarity scoring functions, (3a) hazard identification adaptations that facilitate automatic deletion of irrelevant parts of retrieved cases and integration of all relevant cases, and (3b) risk analysis adaptation that uses the Bayesian approach to integrate both subjective and objective estimates of likelihood to produce a balanced estimation of risk values.

The CBR approach is implemented in a prototype system known as the Safety Knowledge Management System (SKMS). The prototype SKMS was applied on a case study to validate the proposed concepts. The case study is based on a typical work scenario in the construction industry and the case base contained 59 incident cases and 10 risk assessments obtained from different industry sources. The case study shows that based on the relatively small amount of cases, the SKMS is able to retrieve and fully utilize available cases to produce a reasonably thorough risk assessment tree. The case study also demonstrates that a balanced estimation of risk based on both objective and subjective sources can be derived and used to systematically prioritise safety efforts on site.

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

α	Corresponding weight for *common*
β	Corresponding weight *different*
λ	Mean number of incidents in t man-hours worked
$\overline{\lambda}$	Mean or estimated value of λ
$\overline{\lambda}_T$	Estimated mean arrival rate of construction incident for a project
$\overline{\lambda}_j$	Estimated mean arrival rate of construction incident for sub-process j
$f'(\lambda)$	The prior (initial estimated) distribution of the incident rate
$f''(\lambda)$	The posterior (revised) distribution after incorporating incident observations through Bayesian updating
$\overline{\lambda}'$	The prior (initial estimated) $\overline{\lambda}$
$\overline{\lambda}''$	The posterior (revised) $\overline{\lambda}$
κ	The shape parameter for gamma distribution
$1/\upsilon$	The scale parameter for gamma distribution
BE or *B*	Breakdown event
CBR	Case-Based Reasoning
CBRS	Case-Based Reasoning System
CE or *C*	Contact event
common	The number of sub-concepts that are shared between two values (e.g. V_1 and V_2)
C.O.V.	Coefficient of variation

CSQ	Consequence
different	The number of sub-concepts that are not shared between two values (e.g. V_1 and V_2)
E_j	The categorization of the incident event of sub-process *j*
F	Severity category where fatality is involved
GSS	Global similarity score
IE or *I*	Intermediate event
IC	Incident case
k	The normalising constant given by $k = [\int_{-\infty}^{\infty} L(\lambda) f'(\lambda) d\lambda]^{-1}$
L(λ)	The likelihood of observing the incident set assuming $f'(\lambda)$,
L_i	The number of links between the MSCA and V_i
LSS	Local similarity score
L_{Ti}	Total number of links between V_i and the top node
MDL	Man-day lost
mhr	Man-hours worked; number of workers × average number of hours worked
MLCM	Modified Loss Causation Model
MSCA	Most specific common abstraction
OWO	Object-worked-on
p_r	The probability of an incident being reported
S	Severity of consequence
SKMS	Safety Knowledge Management System

SMS	Safety Management System
SS_{GT}	User-specified GSS threshold
SV	Situational Variable
t	Time or time interval
Var(λ)	Variance of the random variable λ
V_i	Value i; a specific value of a type of situational variable
w_{ci}	The weight of the common sub-concept i
w_{dj}	The weight of the sub-concept j that belongs only to one of the values (i.e. not shared)
w_i	The corresponding weight of attribute i
x	The number of incidents recorded in t intervals (of 50,000 man-hrs)
$X(t)$	The total number of construction incidents that had occurred up to time t
X_{ij}	The number of incident occurrences of sub-process j (j = 1 to m) in interval i (i = 1 to n)

Chapter 1

INTRODUCTION

1.1 Poor Safety Performance in the Construction Industry

Safety has always been a perennial problem in the construction industry. In the United States, it was reported that the construction industry accounted for 20% of all occupational fatalities, when they made up only 5% of the United States work force (National Safety Council 1997). In Kuwait, the industry accounts for 42% of all occupational fatalities (Kartam and Bouz 1998) and in Hong Kong the industry accounts for more than a third of all industrial accidents over the last ten years (Tam and Fung 1998). In Singapore, 29% of industrial workers are employed in the construction industry and they accounted for a disproportionate 40% of the industrial accidents (MOM 2001). These studies show that the construction industry has a disturbingly poor safety performance, which translates into much human suffering.

Moreover, the economic cost of an accident is enormous. Based on a study by USA's Center to Protect Workers' Rights (CPWR 1993), the average annual cost of construction accidents (direct and indirect costs) in the United States was estimated to be US$7 billion to US$17 billion. In addition, Everett and Frank (1996) highlighted that the cost of accidents and injuries has risen from a level of 6.5% of construction costs in 1982 to between 8% and 15% during the 1990s.

1.2 The Need for Continual Improvement and Feedback Capabilities

To improve the industry's safety performance, one main strategy would be to ensure continual improvement of safety management systems (SMS) of construction projects. Based on the definition given in British Standard (BS) 8800 (BSI 1996), SMS can be thought of as an interdependent set of preventive measures, which is targeted at maintaining and improving safety performance within an organization. SMS is essentially based on the risk management process (BSI 2000) as illustrated in Figure 1.1, which consists of four interdependent components: hazard identification, risk analysis, risk control selection and risk control implementation and maintenance. In this research, the first two components, i.e. hazard identification and risk analysis are defined as risk assessment, and the first three components, i.e. risk assessment and risk control selection, are defined as safety planning (see Figure 1.1).

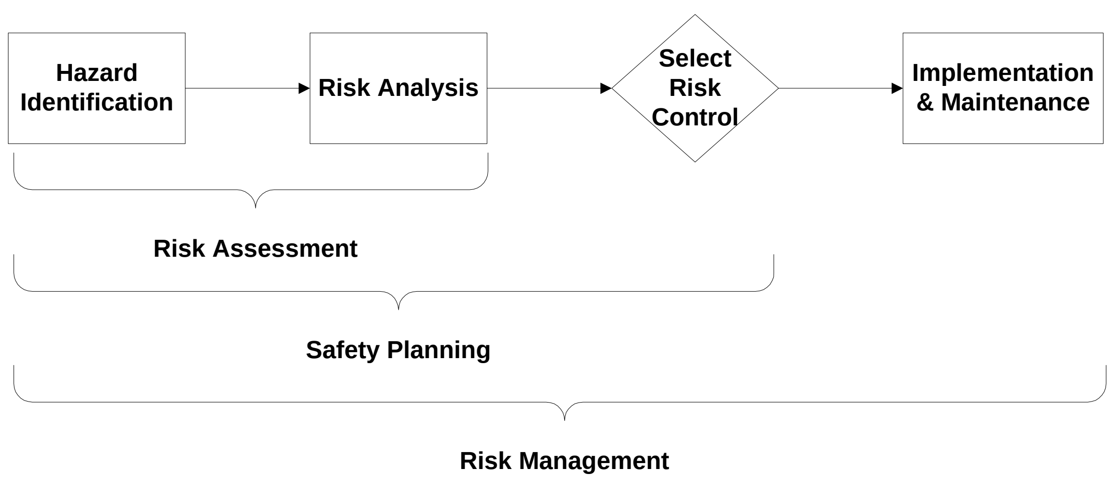


Figure 1.1 Basic Risk Management Model

As shown in Figure 1.2, there are two improvement loops that could be employed to support continual improvement of an SMS. The two loops are facilitated by risk control maintenance and incident investigation respectively. Risk control maintenance is proactive, providing feedback based on pre-planned monitoring and inspection activities, whereas incident investigation is activated only when some kind of physical failure or injury occurs (an incident). Even though the incidents might not result in death or injuries, there would usually be some losses in terms of lost time or damage to property, both of which are also highly undesirable. Thus, incident investigation should not be used as the primary continual improvement measure.

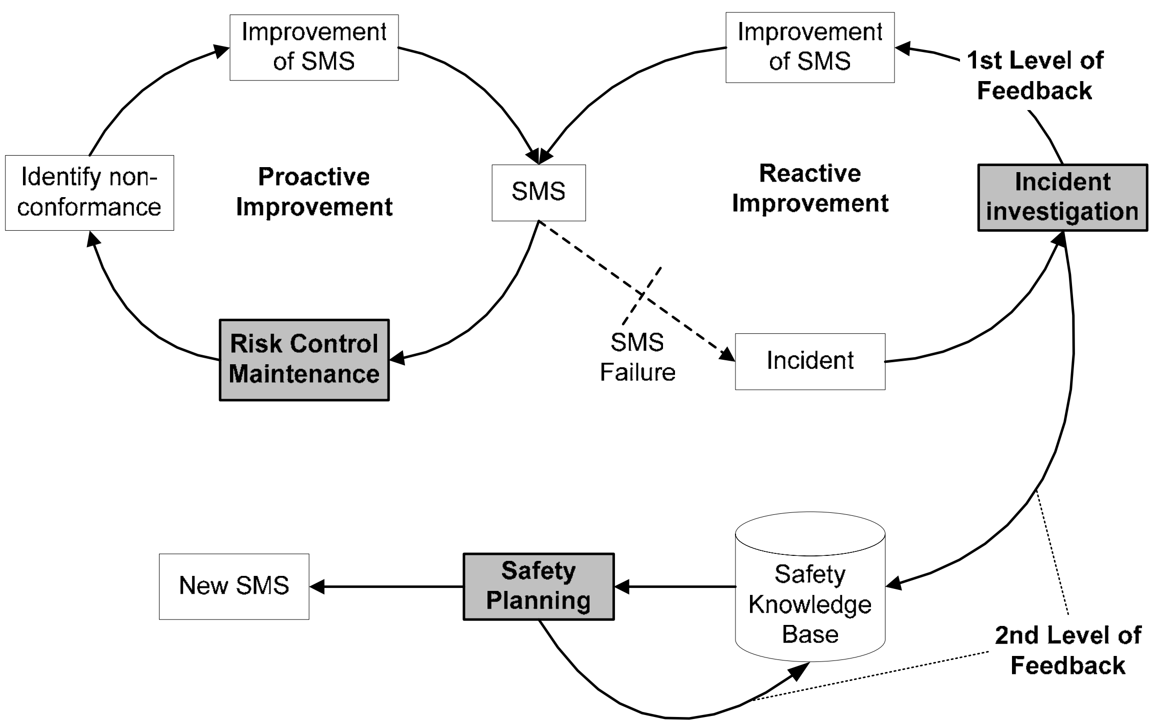


Figure 1.2 Feedback mechanisms to facilitate continual improvement

However, due to the *ex post facto* nature of the information gathered during an investigation, incident investigation information tends to be evidence-based and more convincing. Thus, the information gathered from incident investigations have tremendous value in facilitating improvement of the safety management of construction projects. In order to fully exploit incident investigation information, the incident investigation system should be carefully planned such that it can facilitate feedback at two levels; firstly, feedback to the SMS that had failed (thus causing the incident), and secondly, feedback to the safety planning of future projects (Figure 1.2).

The first level of feedback is within the same project and is more straightforward. The key is to ensure a thorough investigation that identifies the relevant SMS failures so that appropriate improvement to the SMS can be made.

The second level of feedback is not constrained within a single project. It requires the retrieval of relevant safety knowledge from a safety knowledge base, and its adaptation for use in the safety planning of new projects. Safety planning relies heavily on the experience and competency of the safety planning team. The processes of identifying hazards, assigning appropriate level of risk and selecting the most efficient control requires extensive field knowledge and experience. Valuable sources of such experience can be derived from investigation of past incidents. Besides incident investigation information, another possible source of knowledge that should be included in the second level of feedback is the safety plans of past projects. Each safety plan contains possible hazards and proposed risk control measures. Such safety knowledge should also be stored in the knowledge base so that future safety planning teams can retrieve them for adaptation and further improvement.

However, Henderson et al. (2001) identified that most of the companies surveyed (across industries) view incident investigation as a stand alone process that is decoupled from risk management and other proactive measures. Furthermore, the study also showed that there is a lack of computer-based system to manage incident investigation information. With a lack of computer based repositories and linkage between risk assessment and incident investigation, companies as a whole are not able to carry out the two levels of feedback proposed in this research. Furthermore, based on the literature review carried out during this research (to be discussed in chapter 2), it is evident that there is a definite lack of tools to assist companies in realising the two levels of feedback.

1.3 Objectives of Research

This research project aims to provide the necessary framework, concepts and procedures to implement the two levels of feedback effectively and efficiently. The research will develop a prototype system known as the Safety Knowledge Management System (SKMS) and the prototype SKMS will be implemented in a case study to verify the research findings. The SKMS's main purpose is to facilitate the systematic recording and feedback of safety knowledge to improve the effectiveness of safety planning. The key sources of safety knowledge that the SKMS works on include incident cases and past safety plans. Through intelligent retrieval and adaptation of past experiences, the SKMS facilitates systematic organisational learning to prevent recurrence of past mistakes and encourages reuse and improvement of past safety plans.

This research is focused on the hazard identification and risk evaluation portions, i.e. risk assessment, of the risk management process (see Figure 1.1). However, the concepts and methodologies developed in this research will also be the basis for the risk control component of the SKMS. The objectives of this research are as follows:

1. to develop an incident causation model and a common knowledge representation scheme to abstract and capture safety knowledge in incident investigation reports and past safety plans;
2. to propose an intelligent retrieval method that can automatically identify and retrieve relevant past experiences;
3. to propose adaptation strategies to contextualise the retrieved cases for: (a) hazard identification, and (b) risk analysis; and
4. verify the developed and proposed concepts and methodologies through a prototype SKMS, which will be implemented in a case study.

The objectives can be better understood with reference to Figure 1.2. The incident causation model acts as the common underlying framework for both incident investigation and safety planning. It models how and why incidents occur and identifies key knowledge elements that should be captured and utilised during safety planning and incident investigations. The knowledge representation scheme developed based on the incident causation model provides the actual knowledge base structure that will be implemented in the prototype SKMS.

Objectives 2 and 3 focus on developing the retrieval and adaptation components of a proposed SKMS. Through the retrieval and adaptation of past experiences, the second level of feedback can then be achieved. To demonstrate the feasibility of the proposed approach, a prototype SKMS will be developed and verified through a case study (Objective 4).

Based on the literature review (chapter 2) on Information Technology (IT) and Artificial Intelligence (AI), Case Based Reasoning (CBR) (sub-branch of AI) has similar foundational principles as the proposed approach and will be able to facilitate the development of the SKMS. Thus, the key components of the prototype SKMS were developed based on CBR concepts.

1.3.1 Components of the SKMS

To further clarify the objectives, the components of the SKMS are illustrated in the data flow diagrams (DFD) of Figures 1.3 and 1.4. The context level DFD (Figure 1.3) shows that the SKMS interacts with two key interfaces, incident investigation and safety planning. Incident investigation acts as a source of data for the SKMS, where investigation reports are fed into the SKMS. On the other hand, safety planning teams use adapted solutions from the SKMS, and at the same time they also provide the completed safety plans as input to the case base. Thus, safety planning acts both as a sink interface and a source interface.

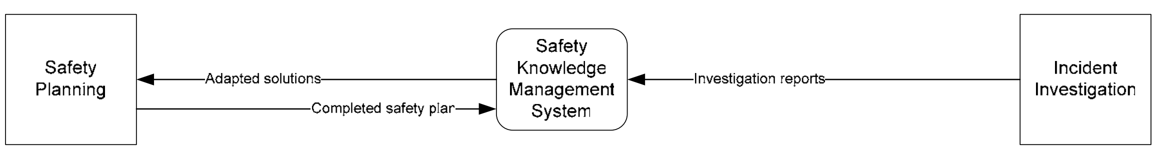


Figure 1.3 Context level data flow diagram of the SKMS

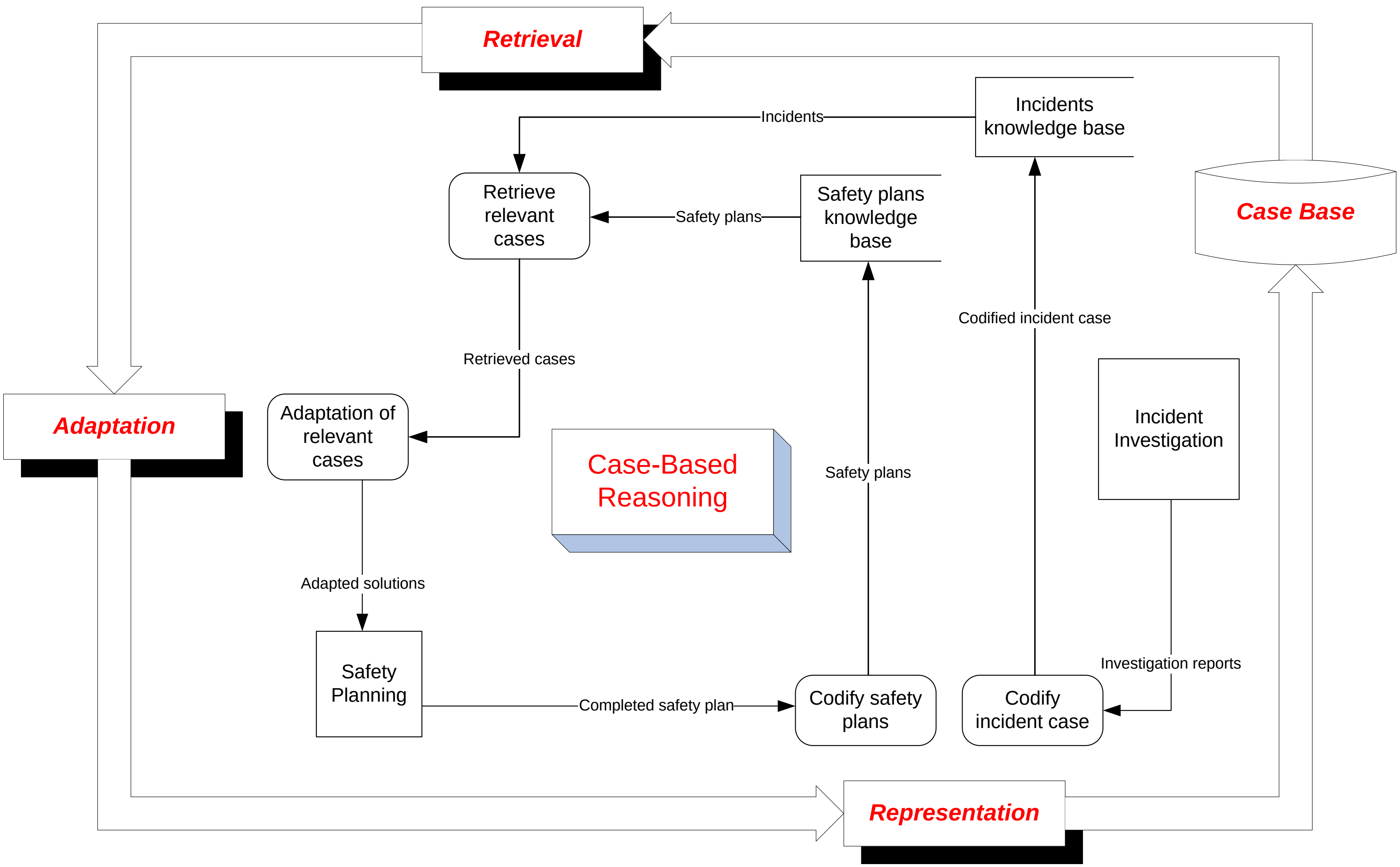


Figure 1.4 Level 1 data flow diagram of the SKMS

Figure 1.4 shows the level 1 DFD of the SKMS, which is an expansion of the context level DFD in Figure 1.3. As can be seen, the SKMS has several inter-connecting codification, retrieval and adaptation processes. These processes correspond to the key components of a CBR approach, which includes knowledge representation, retrieval, and adaptation. The SKMS also contains two key knowledge repositories: the incident knowledge base and the safety plan knowledge base. These knowledge bases correspond to the case base component of a CBR system.

In order for the SKMS to retrieve relevant incident cases and relevant past safety plans, proper codification and indexing of the cases in the knowledge base are important. Each case will have to be abstracted into a manageable codified form, with the appropriate indexes tagged to the case to facilitate retrieval. Besides codification, the retrieval mechanism also requires careful considerations. In order to recall sufficient and appropriate cases the retrieval mechanism must be able to handle inexact matching intelligently. Past cases that are retrieved will need to be adapted to the current context in order for the past knowledge to be more tailored to the present context.

All three key activities of a safety planning process, i.e. hazard identification, risk evaluation, and risk control selection, requires retrieval and adaptation processes. These retrieval and adaptation processes are inter-dependent and similar in principle. Thus, the key SKMS components developed for risk assessment will also be applicable for the risk control component that is not covered in this thesis.

1.4 Scope of Research

As implied in the earlier sections, this research is focused on the construction industry, but the findings and contributions of this research will still be relevant to other industries. Furthermore, despite the broad concepts proposed for risk control selection this research is primarily confined to the area of risk assessment (see Figure 1.1), i.e. hazard identification and risk analysis.

1.5 Research Methodology

Research methodology is made up of two main components: (a) research design, and (b) methods of data collection. As indicated in Figure 1.5, research design is essentially the plan for getting from the research question (or objectives) to the conclusion (Tan 2004). With reference to Figure 1.5, the research designs would then be required for the validation of: (a) the MLCM, and (b) the proposed CBR approach to construction safety risk assessment.

The research design for the validation of the MLCM is more of a case study approach and the method of data collection is essentially analysis of past documents. 140 randomly selected accident investigation reports were obtained from the Ministry of Manpower and the MLCM framework was applied on each of the accident investigation reports to codify and structure key safety information. Each report acts as a case to test the usefulness of the MLCM framework in codifying accident investigation information. Furthermore, the statistics aggregated from the 140 cases also served to validate the effectiveness of the MLCM framework in generating meaningful statistics. It is noted that unlike other research designs involving case studies, this portion of the research used a relatively large number of cases to validate the MLCM. However, the large number of cases is warranted because statistics need to be generated from the cases studies for analysis. Furthermore, it may be argued that the 140 cases is still a relatively small sample (as in most case studies) compared to the wide variety of construction incidents.

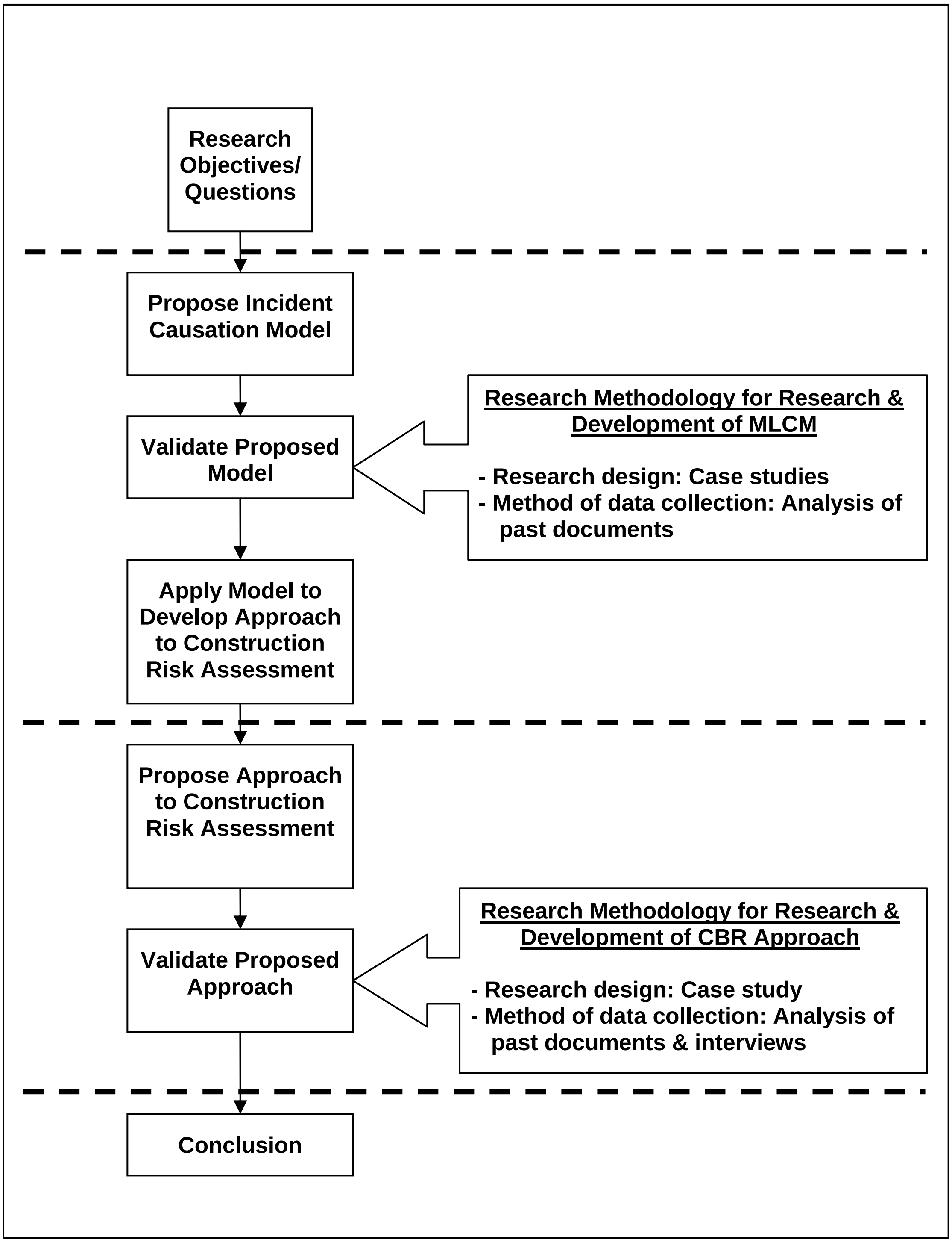


Figure 1.5 Research methodology adopted in this research

As reflected in Figure 1.5, the proposed CBR approach to construction safety risk assessment is also validated through case study. However, unlike the validation of the MLCM, only one in-depth case study was used to validate the approach. For this part of the research, two main types of data were collected: incident cases and risk assessment reports. The incident cases were obtained from the Land Transport Authority (LTA), and the risk assessment reports were obtained from several sources, such as contractors and the LTA. Thus the method of data collection was mainly analysis of past documents, but interviews were also conducted with experienced safety practitioners to ensure completeness of the documents. Typically interviews were conducted to determine likelihood estimates that were missing in some risk assessment reports.

The case study validated the proposed approach by demonstrating how the proposed approach can be applied to develop a reasonably in-depth risk assessment tree for a typical construction activity. Although the data used in the case study was small, the proposed approach was able to be studied in detail to surface the advantages and limitations of the different components of the approach. The case study also showed how the outputs of the approach can be utilised to facilitate prioritisation of risk control efforts.

1.6 Organisation of Thesis

This research will first present the literature review on relevant works in risk assessment and also knowledge management tools in chapter 2. Chapter 3 will present the Modified Loss Causation Model (MLCM), which acts as the broad knowledge framework of the SKMS's knowledge repositories. Chapter 4 will discuss how incident cases and safety plans are codified and indexed to facilitate the retrieval and adaptation processes. Chapter 4 will also show how similarity scores are calculated based on the proposed knowledge codification and indexing methods. Chapter 5 focuses on how the retrieved cases are adapted to assist risk assessment teams in hazard identification and risk analysis. Chapter 6 will present a case study to demonstrate how the concepts presented in earlier chapters are utilised to carry out an actual risk assessment process. Finally, chapter 7 will conclude the thesis and provide suggestions for further research and development.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents a broad review of literature in the areas of risk assessment and computer-based feedback tools. The review is aimed to understand the various types of risk assessment methodologies and assess the strengths and weaknesses of different computer-based feedback tools and technologies. It is noted that subsequent chapters will also present reviews of literature relevant to the content of the chapters.

2.2 Review of Risk Assessment Methodologies

Due to the higher risks involved in industries like the petrochemical and nuclear industries, these industries have developed a large portion of the available risk assessment methodologies (Kumamoto and Henley 1996). However, regardless of the differences in approaches or industries, most, if not all, risk assessment methodologies are similar in terms of basic principles and contain the key components described in Figure 1.1, i.e. hazard identification and risk analysis. Several risk assessment methodologies include risk control selection as a part of risk assessment, but in this research risk control selection is treated as an individual component of safety planning.

Risk assessment methodologies range from quantitative to qualitative types. Quantitative methods usually quantify the risk values based on measurable frequency and severity scales, while qualitative methods uses broad non-measurable categories to indicate the level of risk, frequency and severity. Quantitative methods include methods such as failure modes and effects analysis (FMEA), fault tree analysis (FTA), event tree analysis (ETA) and probabilistic risk analysis (PRA). Qualitative methods include methods like hazard and operability study (HAZOP), what-if analysis, and job hazard analysis (also known as job safety analysis) (Harms-Ringdahl 1993; Ayyub 2003). However, whether the risk assessment method is quantitative or not often depends on whether the risk assessment team utilises a quantitative scale when estimating frequency and severity values. Thus, traditionally qualitative methods can be easily converted into quantitative methods and vice versa.

Some of the more common risk assessment methods will be reviewed in more detail. These risk assessment methods include: (1) fault tree analysis and event tree analysis, (2) FMEA and HAZOP, (3) what-if analysis, and (4) job hazard analysis.

2.2.1 Fault tree analysis and event tree analysis

H. A. Watson of the Bell Telephone Laboratories developed fault tree analysis (FTA) between 1961 and 1962. It is widely used in the safety engineering discipline to deduce the causes of system failures (Livingston et al. 2001) and it has been known to be capable of analysing engineering systems systematically using both quantitative and qualitative approaches (Kumamoto and Henley 1996).

A fault tree model is a graphical model that displays the various logical combinations of component failures that can result in a failure event (also known as top event). There are various types of gates that allow the user to determine the conditions that would allow an event to occur. If the frequencies of the events in a fault tree are available, then the likelihood of the failure event can be calculated objectively. However, even if actual frequencies are not available, subjective estimates of the frequencies can also be given to allow quantitative analysis.

Event tree analysis (ETA) is usually used to study accidental events in a complex engineering system (Kumamoto and Henley 1996). It is based on forward logic, such that it identifies the range of possible subsequent events following an initiating event. These subsequent events focus on the reliability of accident preventing safety systems or failure probability of engineering components. The probability of each event is estimated and the overall reliability of the system can be quantified.

The FTA and ETA are usually conducted hand-in-hand and together they provide a structured risk assessment. Essentially FTA and ETA adopt a "divide and conquer" approach that breaks up the system into hierarchies. Such an approach allows meticulous analysis to be executed.

2.2.2 Failure Modes and Effects Analysis (FMEA), and Hazard and Operability Study (HAZOP)

Failure Modes and Effects (FMEA) and Hazard and Operability Study (HAZOP) are similar risk assessment approaches. Both adopt a systematic component-by-component evaluation of an engineering system, where the effects, probability and severity of a failure of a component are identified (Redmill et al. 1999; Kumamoto and Henley 1996).

In a FMEA the components of a system are listed and the possible failure modes are identified for each component. The analysis also identifies the causes of failures and then the possible effects of the failures. The probability of the failure mode and the severity of the effects are also assessed. Criticality analysis (CA) is then carried out on the FMEA, where criticality is a relative measure of the consequences of a failure mode and its frequency of occurrences. It is noted that the criticality measure is very similar to the definition of risk in most risk assessment methodologies.

Besides the component-based structure, another key characteristic of HAZOP is that it focuses on the use of standardised guide words and process parameters. A HAZOP team will develop the list of guidewords and process parameters prior to the actual study. During the actual study, the effects of the various combinations of the guidewords and process parameters will be analysed. HAZOP is well used in the chemical industry and a detailed study can last two to three weeks.

2.2.3 What-if analysis

What-if analysis uses a creative team brainstorming "what if" questioning approach to the examination of a process to identify potential hazards and their consequences (Crawley and Tyler 2003). Hazards are identified, existing safeguards noted, and qualitative severity and likelihood ratings are assigned to aid in risk screening. Questions that begin with "what-if" are formulated by the risk assessment team members experienced in the process or operation, preferably in advance. The basic steps involved in a what-if analysis are: (1) collect and study background information, (2) conduct preliminary site visits using interviews and “walk-throughs”, (3) design and prepare preliminary “what-ifs” as “seed” questions, (4) facilitate analysis sessions to identify and evaluate hazards/ accident scenarios, and (5) documentation and recommendations.

The what-if analysis is a simple and relatively straightforward risk analysis method that can be readily used in most work situations. However, the flexibility of the method also results in a lack of rigid structure to guide the assessment, and hence the method is not suitable for inexperienced risk assessment teams.

2.2.4 Job hazard analysis (JHA)

The JHA is another widely used technique that is flexible and usually qualitative. The JHA concentrates on the job tasks performed by a person or a group (Harms-Ringdhal 1993). The JHA begins by separating the job into specific and significant job steps. The hazards and possible incidents that can occur are then identified. The risks posed by the hazards and possible incidents are then estimated either qualitatively or quantitatively. Finally, appropriate risk controls are then developed to reduce or eliminate the risks to an acceptable level.

The JHA is a very suitable technique for the construction industry, because the industry is project-based and does not have a fixed working environment or facilities. In contrast to risk assessment methods that focus on systems and their components, JHA provides an appropriate structure for construction risk assessment. Moreover, the construction industry has traditionally used activities for project and work planning purposes. Thus by adopting the JHA approach, the safety plans developed can be more easily integrated into the overall project plans.

In later chapters, the JHA will be used as the basic risk assessment methodology for the SKMS. Useful features like the use of sequential events in ETA and the use of standardised guidewords as in HAZOP will also be incorporated into the JHA method for coding the knowledge in the SKMS. The integrated JHA will be based on the incident causation model developed in this research. Chapter 3 will present the risk assessment methodology in the framework of the proposed incident causation model.

2.3 Review of Relevant Computer-Based Tools for the Construction Industry

Kletz (1994) and Kjellèn (2000) have called for the use of IT to facilitate feedback and learning from past incidents and knowledge. However, based on the study by Henderson et al. (2001), only 15% of the companies (across industries) surveyed use a primarily computer based system to store incident investigation information, and only 24% of the companies use information from incident investigations to conduct their risk assessment or safety planning. Since the construction industry has one of the poorest safety records, it can be inferred that the above mentioned deficiencies are even more severe in this industry. Indeed, based on the literature review conducted during this research, publications on computer based construction safety management tools are rarely found.

The review of construction safety literature from 1994 till 2003 (past ten years) through the Science Citation Index Expanded (Thomson ISI 2003) shows that there had been only two construction safety-related publications that researched on computer-based tools serving some knowledge management purposes. These two research studies were conducted by Kartam (1997) and Hadikusumo and Rowlinson (2002) respectively, and they will be discussed in the following sub-sections.

2.3.1 IKIS-Safety

Kartam (1997) worked on the development of the key concepts for a prototype system known as the integrated knowledge-intensive prototype system for construction safety and health performance control (IKIS-Safety). IKIS-Safety relies heavily on a Database Management System (DBMS) as its knowledge base. The IKIS-Safety was intended to integrate the safety DBMS with a critical path method (CPM) scheduling software, such that for each activity in the scheduling software the relevant safety information in the knowledge base would be tagged onto the activity. Safety activities could also be inserted as an activity in the schedule if the activity is deemed to require visibility.

Kartam's work aimed to provide relevant legislation and experts' recommendations to the project manager through retrieval based on exact matching of indexes like activity code. The IKIS-Safety is a potentially useful tool because project managers are provided with the relevant information for different types of activities on the project schedule. However, the tool is not meant to act as a feedback tool that helps organisations learn from safety knowledge stored in the organisation. Furthermore, the retrieved information will tend to contain precision error (Kjellèn 2000), because the retrieval based on only one exactly matched index may not be able to draw out sufficient relevant information.

2.3.2 Design-for-Safety-Process Tool

Hadikusumo and Rowlinson (2001) attempted to develop a visualization software known as the design-for-safety-process (DFSP) tool. The DFSP tool is meant to facilitate the hazard identification process during the design phase, so that designers can eliminate or minimise the hazards that constructors face during the construction phase. In comparison to 2D plans and drawings, a visualisation tool that is able to represent the construction process dynamically will help designers identify hazards much more effectively.

The DFSP tool has three key components: (1) the virtually real construction model, (2) virtual reality functions, and (3) safety knowledge database. The construction model refers essentially to the construction components of the entity to be built, and the virtual reality functions such as collision detection and terrain following are usually available in commercial visualisation tools. The safety knowledge database in the DFSP tool contains information on construction components/ object types, which acts as the indices for safety knowledge like potential hazards and accident precautions. In this way, users will be alerted of potential hazards and relevant precautions during the simulation of the construction process.

In the context of this thesis, the DFSP tool is similar to the IKIS-Safety in implementation. Even though DFSP tool and IKIS-Safety facilitate safety management and planning, they do not attempt to facilitate the feedback of safety knowledge as proposed in this research. From the angle of retrieval strategy, both employ a DBMS as the safety knowledge base and safety information are retrieved based on exact matching of indexes such as activity code and component type. Due to the shortcoming of traditional database-style retrieval, DFSP tool and IKIS-Safety can easily miss out on relevant hazards or safety information. This point will be further discussed in the next section.

2.4 Tools for Management of Safety Knowledge

The field of knowledge management (KM) arose from the needs of modern companies to acquire, capture, access and reuse knowledge so that they can act intelligently in a sustained manner (Fowler 2000; Wig 1993). A large portion of the KM's development had been initiated by the business-oriented organisations seeking ways to face the challenges of the ever-changing business environment. However, KM concepts can be applied in almost any type of organisation and for a multitude of organisational functions and purposes.

In most KM applications technology is an important aspect. This is because technology serves both as an instrument for knowledge possession and creation and as a possible contributor to the knowledge proliferation and utilisation processes (Hammer and Champy 1993; Davenport and Beers 1995; McQueen and Kock 1996; Davenport 1997; Brown and Duguid 1998). Due to the pivotal role of technology in KM, a large variety of computer based systems have been developed over the years to perform various KM functions. This is in striking contrast to the construction industry.

The KM systems can be based on Artificial Intelligent (AI) tools and Information Technology (IT) like Database Management Systems (DBMS), Knowledge-Based Expert System (KBES), Artificial Neural Network (ANN) tools, and Case-Based Reasoning (CBR) systems (CBRS) (Baets 1998). Each type of technology has its own strengths and weaknesses. Therefore different KM functions, knowledge sources and system environments will call for different types of technology to be applied.

2.4.1 Database Management Systems

A DBMS is defined as software designed to assist in maintaining and utilising large collections of data (Ramankrishnan and Gehrke 2000). Due to its known robustness, efficiency and easy administration, it is widely used in managing data for various purposes, ranging from financial analyses, to maintenance of personnel information in organisations.

However, as mentioned earlier, DBMS as a KM tool has limitations in terms of its retrieval capabilities. Kjellè (2000) broadly described the use of a DBMS to facilitate the abstraction of information from an incident database. He identified two main types of error in carrying out a query on the DBMS. Type I-error concerns with the degree of retrieval, i.e. wanted data not found, and type II-error concerns with the degree of precision, i.e. data obtained are not the wanted data. The reduction of the two types of errors is, to a certain extent, conflicting. Kjellè (1987) found that a skilled user would achieve a higher degree of retrieval through use of free-text searches, as compared to fixed alternatives, but he commented that the higher degree of retrieval might still be at the cost of lower degree of precision.

Watson (1997) further elaborated the retrieval problems in DBMS. He highlighted that DBMS is not able to handle fuzzy matches well, and a lot of real world problems require such fuzzy matching capability. A DBMS mainly uses keyword search, wildcards and logical operators to handle ambiguous search. These DBMS search methods are fairly efficient in handling well-defined and straightforward problems, but in complex searches, such as in the case of the safety planning tasks that the SKMS aims to facilitate, a more flexible and intelligent search method is needed so that similar concepts or relevant hazards which may be relevant to a specific case are reviewed.

2.4.2 Knowledge-Based Expert System

Knowledge-Based Expert System (KBES) is one of the earliest forms of AI tools. It is characterised by its use of large bodies of domain knowledge, facts and procedures gleaned from human experts that have proved useful for solving typical problems in their domain (Dym and Levitt 1991). The knowledge captured in a classical KBES is usually in the form of IF-Then rules abstracted from experts. By capturing the IF-Then rules and relevant facts into the knowledge base, an inference engine can be used to help non-experts to gain access to invaluable expert knowledge, and hence improve the quality and efficiency of work.

The chemical industry has made numerous attempts in developing KBES to improve and automate hazard identification (Catino and Ungar 1995; Chae and Yoon 1994; Suh et al. 1997a, 1997b; Vaidhyanathan and Venkatasubramanisn 1996; Weatherill and Cameron 1989). The impetus for these KBES is mainly to shorten the time taken for detailed hazard analyses. As in most KBES, these systems are usually based on experts' context-specific causation models and a knowledge representation scheme that is unique to the scope of the research. For instance, the automatic hazard analyser (AHA) developed by Kang et al. (1999) is based on several knowledge bases that contain information like the spatial arrangement of process units, the connective relation among process units and hazardous characteristics of materials. Furthermore, the inference process is linked to the process units of the plant. Thus the knowledge and inference process of these KBES is only applicable to the intended area of application and is not easily adapted to other applications.

KBES is also known for its difficulty in implementation. The acquisition of expert knowledge is often tedious and labour intensive (Holland 1986). More often than not, KBES developers and researchers are bogged down by the knowledge acquisition phase, because knowledge engineers need to spend a large amount of time to elicit and abstract expert knowledge from tacit form to explicit form and finally to rules suitable for use in KBES.

Furthermore, the level of difficulty is magnified due to the characteristics of the construction industry. Even though there are a large number of KBES developed in the construction industry Mohan (1990) and Li (1996) noted that the changing environment of the construction industry makes it unsuitable for systems or tools to be bound by the rules in most KBES. That is because frequent changes in the environment may result in errors in the inference process and time-consuming maintenance of the rules. Tah and Howes (1998) even felt that despite the large number of KBES developed in the construction industry, they have failed to make an impact.

Moreover, KBES is also unsuitable for the SKMS because it requires clear rules that relate cause and effect. However, there is no incident causation theory that can reliably relate causes and effects of construction incidents. Thus, the rules developed may not be robust enough to allow a KBES to facilitate the safety planning process.

2.4.3 Artificial Neural Networks

Artificial Neural Networks (ANN) has the ability to abstract information from a pool of incomplete experiences, generalise and apply the learned knowledge in new situations (Kasabov 1999). ANN is based on an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the human brain processes information (Russel and Norvig 1995). Each ANN consists of a number of nodes or units connected by links. Each link has a numeric weight associated with it, and these weights can be updated to align the network with the inputs that it receives from input nodes.

ANN had been widely used to analyse complex information to identify patterns and produce networks that has the ability to estimate likely output relevant to construction projects. For example, Chua et al. (1997) utilised the ANN approach to model construction budget performance. The network developed by Chua et al. (1997) can be used by construction companies to evaluate management strategies and make resource allocation decisions. Another study by Hegazy and Ayed (1998) also used the ANN approach to estimate the cost of highway projects. In their study, the advantage of ANN is that it is able to handle the complexity in cost estimation, which contrasted with the limitations of regression models.

However, ANN has several drawbacks. Firstly, the training process of the ANN is not transparent, making it hard for users to trace and understand the rationale of the output (Hegazt et al. 1994). In this way, even though knowledge is captured in the form of the ANN, it is not easily transferred to humans. In the case of the SKMS, comprehension of the output is very important because the safety planning team will need to be able to understand the safety plans in order to implement the safety plans effectively. Secondly, the ANN training process requires much trial and error, and the learning algorithm needs to be carefully selected. Hence, the training of the ANN also requires certain amount of expertise. Thirdly, ANN is more suited for quantitative data, rather than symbolic and qualitative data, which make up most of the knowledge used by the SKMS. Even though it is possible to quantify qualitative data by imposing a linear or even fuzzy scale, this will introduce potential bias into the network. Hence, the ANN will not be a suitable technology for the SKMS.

2.4.4 Case-Based Reasoning Systems

Case based reasoning (CBR) is a relatively new branch of AI, but in recent years there had been an increase in uptake of CBR concepts (Watson 1997). CBR has its root in psychological theory of human reasoning, which has the intuitive paradigm that humans solve new problems by recalling past experiences (Mount and Liao 2001). Referring to Figure 2.1, a CBR system (CBRS) has three key processes: (1) case representation and indexing, (2) retrieval of cases, and (3) case utilisation and adaptation (Kolodner 1993). Case representation and indexing is the process of codifying the lessons that a case teaches and the context in which the case can teach its lessons. Retrieval of cases is the process of searching and determining relevance of past cases in the case base. Case utilisation and adaptation refers to the process of making changes to the retrieved cases to suit the new situation and harnessing the retrieved knowledge to meet the purpose of the users. However, currently a large number of CBRS do not have an adaptation module (Cunningham and Bonzano 1999), and the reason for this is that the context and purpose of the users may be too complex for simple adaptation strategies.

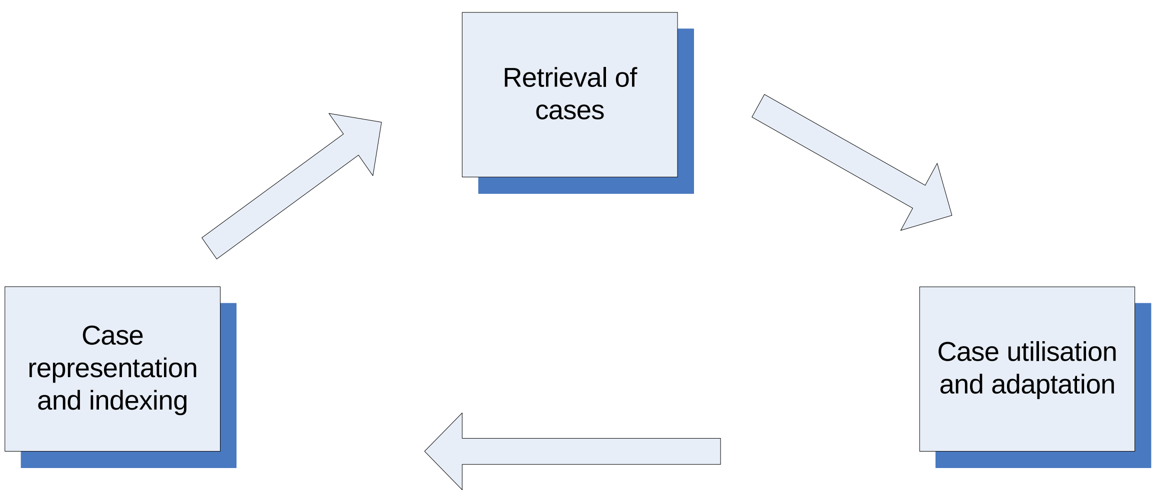


Figure 2.1 Case-based reasoning process

CBR is usually applied in “weak theory” domains where the key causal relationship and interactions is uncertain or too complex to be modelled effectively. Early developments in CBRS resulted in a wide range of tools such as Cyrus (Kolodner 1984), MEDIATOR (Simpson 1985), CHEF (Hammond 1986a), PERSUADER (Sycara 1987), HYPO (Ashley 1988), CASEY (Koton 1989) and JULIA (Hinrichs 1992). These early tools are applied in domains like law, cooking, legal disputes, labour mediation and medicine, where a large amount of judgement and intuition is needed to make decisions and the available information are usually not numerical.

CBR concepts had also been successfully applied in different engineering fields, for example Chua and Li (2001) applied CBR concepts to the construction contract bidding process. They adopted CBR concepts because CBRS is able to simultaneously process a large number of highly interrelated variables to arrive at a decision. Liao et al. (2000) also investigated the usefulness of a CBRS by developing a diagnostic system that identifies failure mechanisms of engineering components. Their study shows that the CBRS outperformed rule-based (KBES) and ANN systems. CBRS had also been applied in various areas of the construction industry, for example, structural design (Gero 1990; Marha and Garza 1996), industrial building design (Börner 1995), office design (Pearce et al. 1992), building regulation interpretation (Yang and Robertson 1994), contractor prequalification (Ng et al. 1998) and material selection (Dutton and Maun 1997).

Despite its usefulness, Ong (2002) noted that CBR is not advantageous if the domain is well-understood and well-structured, where rules can be easily defined. Domains and tasks that do not require heuristics, but instead needs specific models for causal reasoning are also not suitable for a CBRS. Furthermore, there should also be sufficient cases, because the usefulness of a CBRS is dependent on the quality and availability of cases. Thus, the suitability of a CBRS for a particular problem domain has to be carefully assessed prior to development.

In comparison to the other discussed technologies and in view of the context of the SKMS, a CBRS appears to be the most feasible form of technology. This is because the sources of safety knowledge, which includes incident cases and past safety plans, are naturally episodic and found in the form of cases, where each case contains information on the hazards and events related to a specific work situation. These cases are also readily available because for more serious cases, government agencies will be required to investigate the incidents. Besides, it is also becoming a norm for construction companies to have procedures in place to ensure that safety plans are produced prior to actual commencement of work, and that most incidents are recorded and investigated.

A CBRS also has intelligent and fuzzy retrieval capabilities, which allows the past knowledge to be retrieved even if the new situation is not exactly the same as the case's situation. This capability is important because most construction work situations are unique and retrieval based on exact matching will result in relevant knowledge to be overlooked. Thus, comparing with a classical DBMS, CBRS will have significant advantages.

In comparison to KBES, CBRS also proves to be a better choice because, unlike KBES, it is meant to function in "weak theory" domains. Furthermore, CBRS is also relatively easy to maintain, because cases can be easily deleted or added without affecting the performance of the system. Thus, a CBRS is more suited to the dynamic nature of the construction industry.

In terms of comprehensibility of output, a CBRS has significant merits as compared to an ANN. When past cases are retrieved, a user can easily comprehend the rationale of the retrieval based on the matching of the indeces of the case and the specific knowledge enclosed in the case. This advantage places CBRS as a more desirable choice than ANN. Besides this, the safety knowledge available to the SKMS is qualitative and symbolic. Hence, ANN will not be able to easily manage the qualitative data effectively and reliably. This is in contrast to CBRS's ability to handle qualitative and symbolic data.

2.5 Conclusions

The literature review shows that there are many possible risk assessment methodologies and the methodologies range from quantitative to qualitative methods. These methodologies are essentially the same in terms of basic principles. Since most construction projects use activities or tasks as the main planning variable, the job hazard analysis (JHA) was deemed to be a suitable technique to be applied in the SKMS. However, the relevant characteristics of other risk assessment methodologies will also be incorporated into the JHA.

Besides that, based on the literature review it can be concluded that there is a lack of efficient and effective computer based tools to facilitate the reuse of safety knowledge in construction safety planning. Only two computer-based tools are considered related to the context of this research, but both tools do not directly facilitate feedback of safety knowledge. Furthermore, the two computer-based tools employ DBMS style retrieval to retrieve safety knowledge. Such an approach is deemed to be inapt for the complex functions of the SKMS.

The review of KM literature shows that commonly used artificial intelligence tools and information technology used are database management system, knowledge-based expert system, artificial neural network and case-based reasoning system. Each type of technology had has its own distinctive strengths and weaknesses. A comparison of the needs of the SKMS to the characteristics of each type of technology shows that CBRS is the most appropriate technology for developing the framework that will facilitate feedback of past safety knowledge to improve construction safety planning

Chapter 3

THE MODIFIED LOSS CAUSATION MODEL

3.1 Introduction

This chapter presents the Modified Loss Causation Model (MLCM), which acts as the underlying framework for the SKMS. The chapter will first review relevant works in incident causation models and then the MLCM will be introduced. Subsequently, the application of the MLCM in incident investigation (first level of feedback) and safety planning (second level of feedback) will be discussed.

3.2 Relevant Works

Following the seminal work by Heinrich (1939), numerous other incident causation models (also known as accident causation models) have been developed. These incident causation models differ in many fundamental ways and may be classified based on their area of application, general structure and key characteristics. The relevant incident causation models are categorized into three general categories: energy transfer models, individual specific models, and systemic models.

The energy transfer models, as the name implies, focus on the transmission of uncontrolled energy from the source to the victim. The energy transfer model developed by Haddon (1980) has much relevance to this study. Haddon developed the model and proposed ten basic prevention strategies based on the points of intervention, i.e., the energy source, the barriers (between victim and energy source) and the victim. The model is very useful in categorizing the types of preventive measures, but by itself the model does not provide a suitable feedback-oriented framework for incident investigation and safety planning.

Individual specific models are models that place emphasis on individuals that contributed to the incident in a direct way. These models identify the causes and effects of erroneous acts by individuals (usually front-line workers). They (for example, Kerr 1950, Kerr 1957, and Hinze 1997) usually focus on psychological and behavioral aspects of humans. However, these models do not emphasise the role of the organization and the safety management system (SMS). Therefore, individual specific models do not explicitly facilitate continual improvement of SMS.

Systemic models refer to models that highlight the role of the organization and its systems in the causation of the incident. Henderson et al. (2001) regarded a system-based approach as one of the requirements of a successful incident investigation, and this view is reflective of the current underlying concept of incident causation. Numerous systemic models have been developed over the years, for example, Management Oversight Risk Tree (MORT) (Johnson 1980), contributing factors in accident causation (CFAC) model (Sanders and Shaw 1988), pathogen model (Reason 1990), Whittington et al.'s (1992) model, Loss Causation Model (Bird and Germain 1996), accident root causes tracing model (ARCTM) (Abdelhamid and Everett 2000) and Constraint-Response Model (Suraji et al. 2001). In these models the organization as a whole plays an important part in the causation of an incident. However, these causes are generally latent (Reason 1993), that is, these causes or failures reside in the organization and only when local triggers (or immediate causes) arise, incidents may occur. These models also implicitly or explicitly reinforce the concept of multiple-causation, where the cause of an incident does not lie in one line of causation, but often branch out into multiple levels of factors. Due to the emphasis on contributions of the organization, systemic models provide the basic framework for the development of the proposed incident causation model.

3.3 The MLCM

The proposed incident causation model, the Modified Loss Causation Model (MLCM) depicted in Figure 3.1, is primarily based on the Loss Causation Model (LCM) by Bird and Germain (1996) and useful features of various incident causation models that are reviewed. The MLCM also incorporated insights obtained from the evaluation of 140 fatal accident cases. The 140 fatal accident cases will be discussed in more detail in the next section.

The LCM, which is the basis for the MLCM, has several useful characteristics. One of the main merits of the model is that it promotes proactive thinking (Covey 1989) on the part of management, which in turn facilitates feedback. In particular, the model identifies “Lack of control” as the fundamental source of incident occurrence, hence prompting investigators to end each incident investigation with an examination of the Safety Management System (SMS). The model, therefore, encourages organizations to accept the responsibility to respond to incidents and not blame it on individuals or physical conditions. In this way, each time an incident occurs the planned SMS will be reviewed and compared with the causation factors identified to determine whether there is a lack of measures to control the occurrence of the causation factor, an inadequacy in the planned risk control, or whether the planned measures were inadequately executed. In this way, systemic actions can then be implemented to remove flaws in management system and organisational culture.

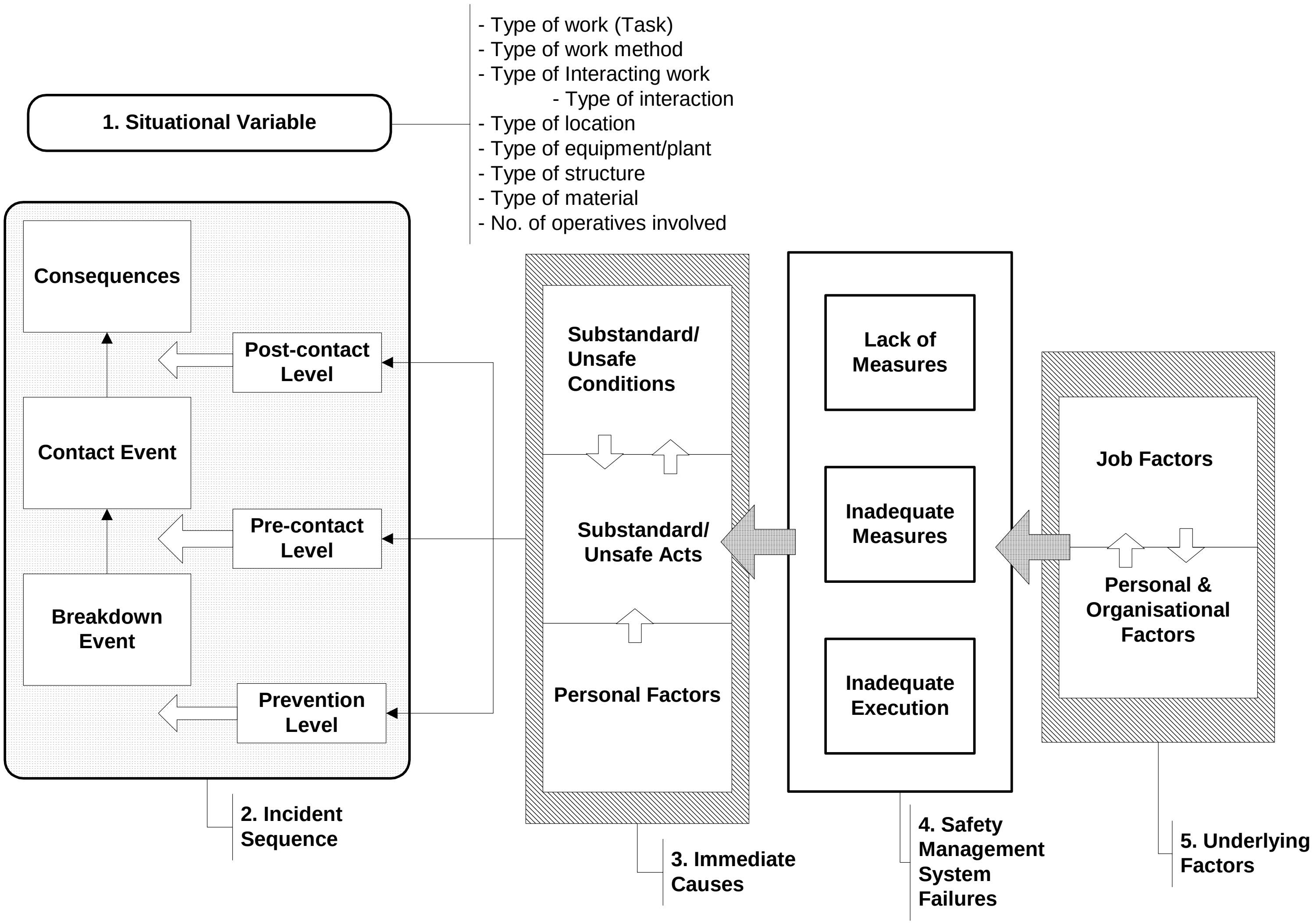


Figure 3.1 The Modified Loss Causation Model

Another useful characteristic of the LCM is that it clearly identifies and distinguishes immediate causes and underlying factors (also known as basic causes). The immediate causes are the triggers that directly lead to the incident sequence, whereas the underlying factors are factors that contribute indirectly to the occurrence of the immediate causes. The underlying factors are usually hidden in the organization and are hard to detect. They are also often contributory in nature and their determination may have to depend on investigators' subjective judgment, but their clear identification can usually lead to more significant improvements in safety performance. In the LCM, immediate causes are classified into substandard/unsafe conditions and substandard/unsafe acts, which refer to the respective physical conditions and human behaviours that do not meet safety requirements and can directly cause incident occurrences. Underlying factors are also categorized into two sub-categories, personal factors and job (or system) factors. Personal factors are defined as factors related to individual's capability, knowledge, skills, attitude and motivation. On the other hand, job factors refer to factors related to work or task definition and execution, for example inadequate leadership and/or supervision, inadequate engineering and inadequate work standards.

Modifications to the LCM have been made to achieve the objectives of this research and the modified version of the LCM, the MLCM, is presented in Figure 3.1. The MLCM is composed of five main components, namely: situational variables, incident sequence, immediate causes, SMS failures and underlying factors (Goh and Chua 2002, Chua and Goh 2004b).

The component "situational variables" has been included into the model because for each chain of incident causation there is a need to identify the critical characteristics of the context or situation in which the incident occurred. In this way, the information and learning points derived from the incident investigations can be more easily identified and applied to similar contexts or situations across different construction projects. This is especially valuable to an industry which is project-based in nature so that the experience gleaned from one project can be transferred to other projects. Moreover, these situational variables serve as indices for maintaining the incident investigation information in the safety knowledge database and for retrieval of related incident experience for safety planning.

The situational variables can also act as stratifying or categorical variables during data analysis of incident statistics, so as to allow meaningful comparison of statistical results. For instance, the number of incidents can be stratified based on the type of work executed prior to the incident. Statistics based on the type of work will provide insights on the contribution of different work types to the occurrence of incidents. Some of the possible categories of situational variables are listed in Figure 3.1.

The second component of the MLCM, the incident sequence, is based on Haddon's (1980) energy transfer model and is made up of the breakdown event, contact event and consequences. The breakdown event is defined as an initiating point of loss of control of a source of energy or substance that, without an intervening event, will lead to the occurrence of a contact event. In contrast, contact event is an event where the victim comes into contact with the source of energy or substance. The consequences refer to the undesirable effects of the incident, for example, property loss, number of man-days lost and type of injuries. It is beneficial to define incidents based on the incident sequence structure so that causal factors and safety measures can be classified systematically into three levels of intervention and causation, namely, post-contact, pre-contact and prevention levels (refer to Figure 3.1).

During incident investigation, the incident sequence structure of the MLCM will prompt investigators to broaden their scope of investigation, and not end an investigation prematurely. For example, when a worker loses balance and falls off the edge of a building, an investigator could easily state that the "main cause" of the accident is due to the fact that the worker was not wearing safety belt. Even if the underlying factors and the SMS failures that had contributed to the contact event (falling from height, subsequently striking the ground) were identified, the knowledge would only prevent the recurrence of the contact event, but not the breakdown event (loss of balance, in this instance). To better improve the SMS, the factors that contributed to the occurrence of the breakdown event, contact event and the consequences of the incident should be identified. Similarly, safety planning should also be based on the three levels of intervention to identify sufficient measures to prevent and mitigate the consequences of the breakdown events and contact events.

The third and fifth components of the MLCM are the immediate causes and underlying factors, respectively. Unlike the LCM, which only has personal factors in the underlying factors, the MLCM include personal factors in both immediate causes and underlying factors. This is to prevent difficulties in classification of personal factors, particularly in cases where the identified personal factors do not fit the definition of underlying factors. For instance, when a worker committed a substandard act by climbing up the bracings of an access scaffold, and it was identified that the personal factor that led to this substandard act is improper motivation to save time and effort. Under the LCM the personal factor will have to be classified as an underlying factor. However, this classification will not fit the definition of underlying factors, which are organisational and/or contributory in nature. In the MLCM, personal factors that lead to substandard acts of front line operatives are separated from personal factors that influence SMS failures and job factors. In this example, an underlying personal factor could be the lack of experience of the safety planning team to develop adequate measures (SMS failure) to deal with the improper motivation (immediate personal factor) that led to the substandard act. Furthermore, organisational factors are also included in the underlying factors to take into account its effects on SMS. Organisational factors include factors like poor safety culture and inappropriate organisational structure.

The fourth component of the MLCM is SMS failures, which can be further classified into: lack of measures, inadequate execution and inadequate measure. This component is similar to the "lack of control" component in the LCM, but in the MLCM the SMS failure is identified prior to the underlying factors, which is the reverse of the LCM approach. After attempts to apply the LCM approach in this study, it was realised that the classification of SMS failures is often a prerequisite to the identification of job factors, thus leading to the order proposed in the MLCM. This investigation approach will be elaborated subsequently.

Another feature of the MLCM is the explicit identification of directions of influence and causation between the various types of factors. In the MLCM (Figure 3.1), substandard acts can be influenced by substandard conditions and vice versa; substandard acts can also be caused by personal factors. In the model, SMS failures can lead to the occurrence of immediate causes of an incident, and underlying factors are deemed to contribute to the failure of the SMS. Within underlying factors, job factors influence personal and organisational factors, and vice versa. These possible directions of influence and causation form part of the MLCM framework for incident investigation and safety planning.

3.4 Application in Incident Investigation

Incident investigation is a key process in the first level of feedback, i.e. feedback to the Safety Management System (SMS) that failed (see Figure 1.2). The MLCM can be utilized to facilitate and improve the investigation process. The following sub-sections show how the MLCM can be used to guide incident investigation, provide a structure for the incident investigation information and also codify the information to facilitate statistical analysis and storage in the knowledge base.

3.4.1 MLCM Investigation Approach

In order to facilitate the first level of feedback, the MLCM is designed to guide incident investigation to uncover SMS failures and underlying factors. Based on the MLCM, the flow chart depicted in Figure 3.2 is developed to illustrate the MLCM approach to identification of SMS failures and underlying factors.

The first step in the investigation would be to identify the situational variables, incident sequence and immediate causes. This initial information is usually the focus of current investigation approaches. However, in order to ensure improvement in the SMS, root causes have to be uncovered. Thus, in the MLCM approach, the relevant

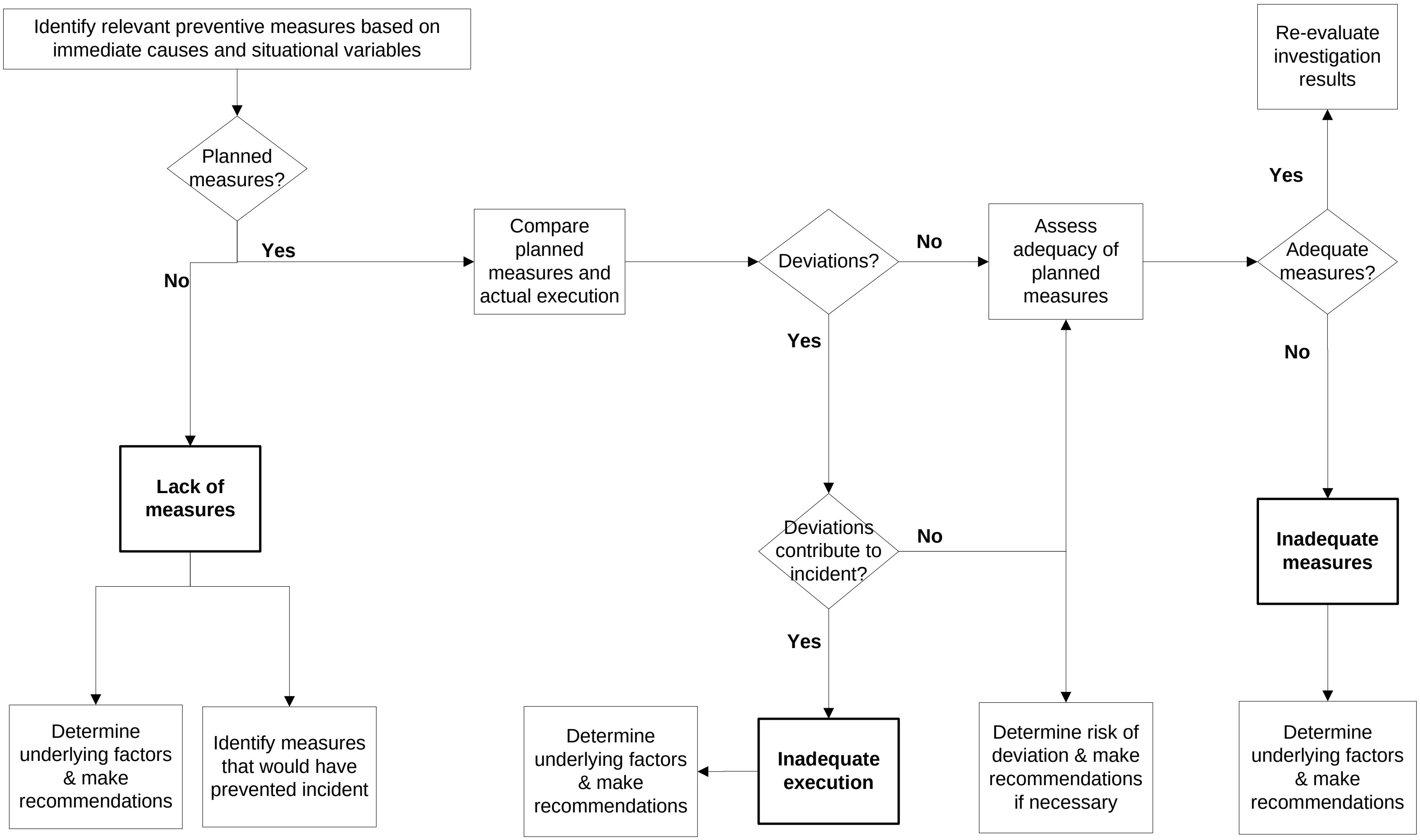


Figure 3.2 The MLCM Investigation Flow Chart

safety measures that could have prevented the immediate causes are identified based on the situational variables, incident sequence and immediate causes. If no relevant measures exist, then there is a failure in the SMS of the "lack of measures" type and the investigation will focus on identifying the underlying factors that lead to this failure, and subsequently propose appropriate safety measures to prevent a recurrence. If relevant safety measures exist, the execution of the measures will be evaluated based on the planned procedures.

When there are no deviations from the planned procedures, the adequacy of the planned measures is next evaluated. For any measures that are inadequate, the investigators will determine the underlying factors that led to the SMS failure and recommend appropriate ramifications. Otherwise, the investigators will re-evaluate their assessment of the incident, since from a proactive mindset few or no incidents are unpreventable.

If the deviations from the planned procedures had contributed or caused the incident, then there is an inadequate execution of the planned measures so that the underlying factors causing the SMS failure have to be detected, and the rectifications made. On the other hand, when the deviations do not contribute to the incident directly, there is a need to assess the adequacy of the planned measures and the risk posed by the deviation.

3.4.2 Structure for Incident Investigation Information

In this sub-section, a case study will be presented to illustrate the usefulness of the MLCM in providing a structure for the information derived from an investigation. The case is based on an actual fatal incident investigated by the Singapore's Ministry of

Manpower, Occupational Safety Department. The incident shown schematically in Figure 3.3, occurred during a lifting operation, which involved the use of a crawler crane. The crawler crane driver was requested to lift a bundle of rebars to the fourth level of a building under construction. During the lifting operation the victim was doing some general work on the fourth level of the building under construction; the location was near to where the rebars were to be placed. When the boom angle of the crane reached approximately 60°, the overload alarm sounded and the crane operator lowered the load quickly. In the process of releasing the hoist rope, the crawler crane tilted and the boom hit the access scaffold. As a result, the bundle of rebars fell onto the victim, who was killed on the spot.

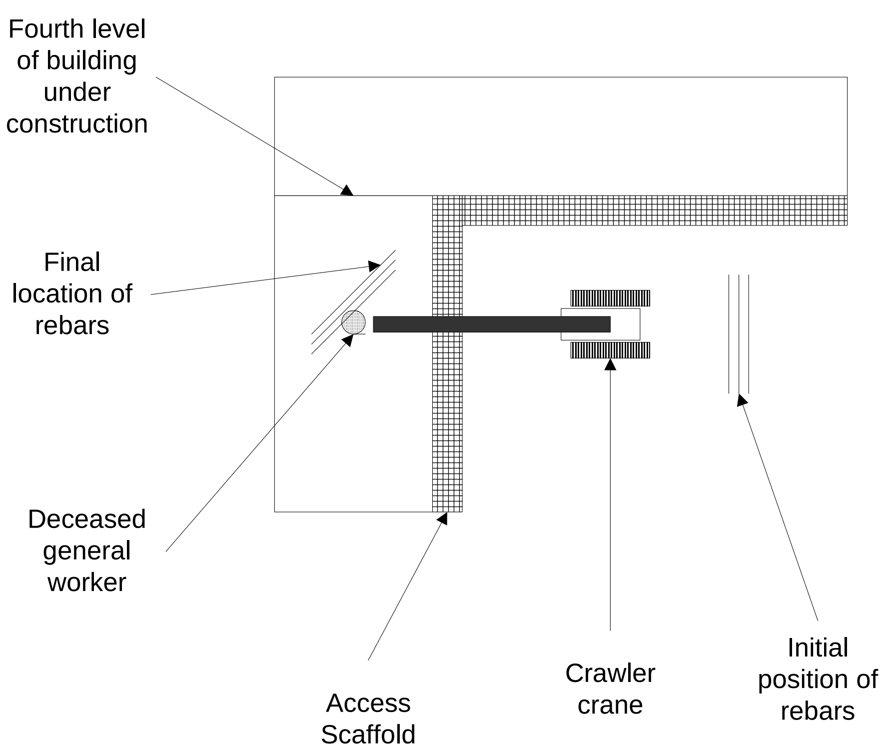


Figure 3.3 Schematics of incident case study

The information from the incident investigation can be structured based on the MLCM as depicted in Figure 3.4. The key situational variables of the incident are the type of work (lifting operation), type of interacting work (general work), type of equipment/plant (crawler crane), type of structure (access scaffold (nearby)) and type of materials (rebars). The death of the general worker (consequence) is due to him being struck by the falling rebars (contact event), which had fallen due to the crawler crane losing its stability (breakdown event). The crawler crane and the nearby access scaffold were also damaged due to the breakdown event (consequence).

The investigation report revealed no information at the pre-contact and post-contact levels. There had been no investigation on what could have prevented the contact event to occur (pre-contact level measures) even after the breakdown event had occurred, and what was done to deal with the emergency (post-contact level measures) after the contact event had taken place. However, the investigator did identify causal factors at the prevention level. The substandard act was the overloading of the crawler crane, and the personal factor leading to the substandard act was the crane operator's underestimation of the load. The investigation revealed a lack of explicit measures in the SMS to prevent the occurrence of the immediate causes. If there had been lifting supervisors appointed and proper measures to ensure that the weight of the load was clearly communicated and determined prior to the lift, the incident could have been prevented. However, the investigation did not attempt to determine the job, personal or organisational factors for the lack of measures, which would constitute the underlying factors for the SMS failures.

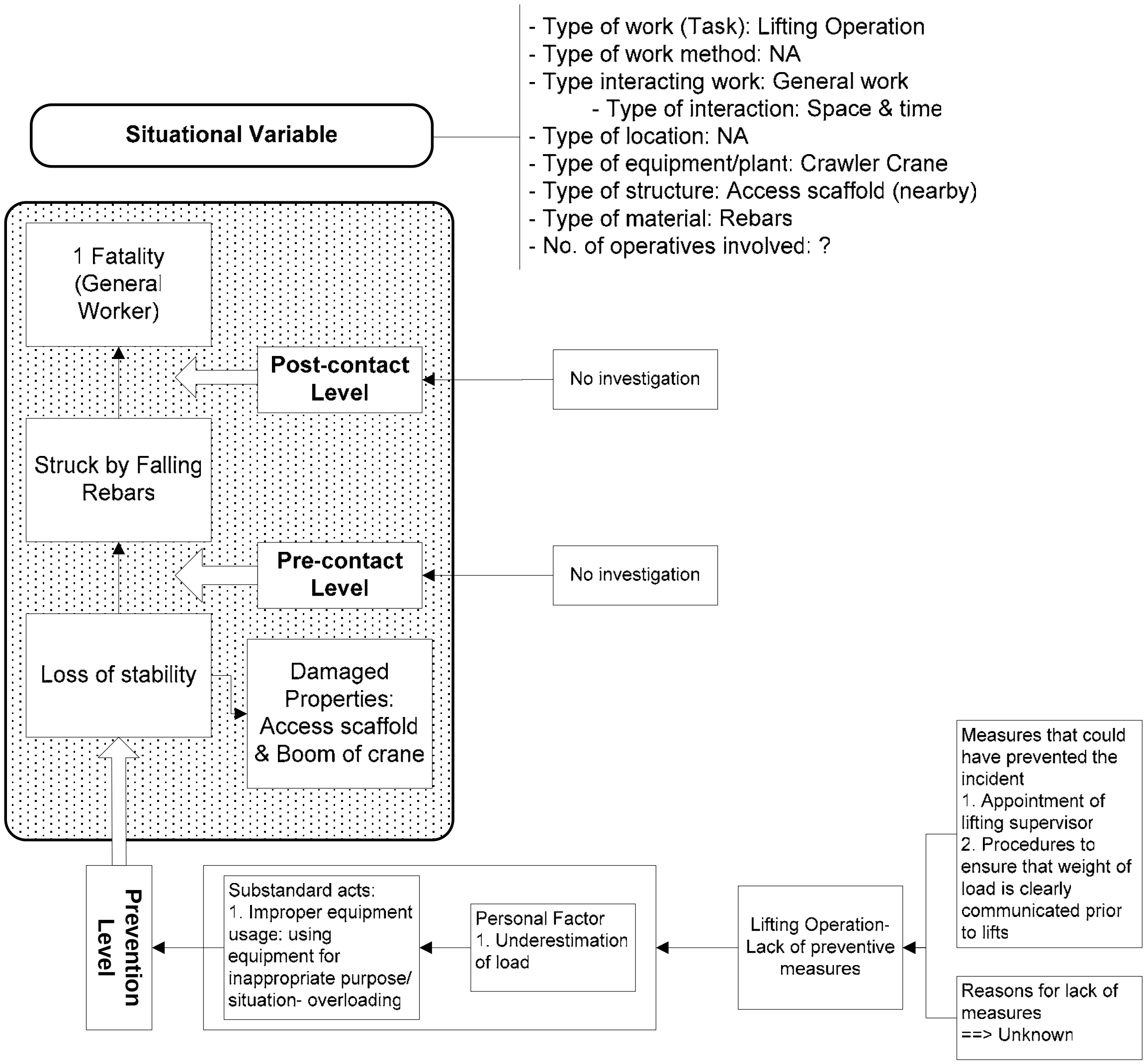


Figure 3.4 Application of MLCM to structure incident investigation information

From the case study, it can be seen that if the MLCM investigation approach had been adopted and applied during the investigation, the level one feedback (feedback to improve the SMS that failed) would be achieved. Even though the investigator did identify the type of SMS failure (lack of measures) at the prevention level, the investigator did not identify the underlying factors that contributed to the failure of the SMS, which is essential to ensure improvement of the SMS. Identifying deep-rooted job, personal and organisational factors is difficult, but it is only through their identification that effective strategies can be implemented to ensure improvement of the SMS.

Furthermore, if the investigation had been carried out based on the MLCM approach, pre-contact and post-contact measures could then be identified or improved. Based on the preceding case, for example, if there had been a lack of measures at the pre-contact level, the investigator could recommend danger areas to be clearly identified or even barricaded during lifting operations. In this way, the contact event of worker being struck by falling loads could be prevented.

3.4.3 MLCM Taxonomy

To facilitate storage of safety knowledge structure based on the MLCM, a set of taxonomy has been developed from the literature review and the study of 140 incident investigation cases obtained from the Ministry of Manpower (MOM), Occupational Safety Department (OSD) (Chua and Goh 2004a). The 140 incident cases have at least one fatality each. Fatal cases were chosen because these cases involve a more thorough investigation and hence contain more details for analysis. It is noted that even though only fatal cases were used, the same set of taxonomy can also be applied to non-fatal and non-injury incidents. Furthermore, the MLCM taxonomy is designed to be generic, so as to cover various types of construction situations and factors. This generic set of taxonomy is then customised to facilitate knowledge representation in the SKMS (see chapter 4).

A thorough review on existing incident taxonomies and classifications was conducted with the aim to identify suitable categorisations for each of the components in the MLCM. Most of the available taxonomies lack a strong underlying incident causation model (Hinze et al. 1998; Kartam and Bouz 1998; Feyer and Williamson 1991; Sawacha et al. 1999), making the logic structure of these taxonomies harder to grasp. Bird and Germain (1996), and Gordon (1998) developed taxonomies that were relatively comprehensive, but not tailored to the context of construction industry. As a result there are difficulties in classifying the incident information in some categories, especially the job-related factors which differ from the construction context. Furthermore, some parts of the taxonomies were split into very detailed factors without sub-categorisations. Categorising incident investigation information into detailed factors allows more specific knowledge to be gained, but more often than not, such a categorisation approach causes data to be too sparse and hence resulting in difficulty in retrieval and reuse of past safety knowledge. Whittington et al. (1992) proposed a set of taxonomy based on the construction industry; this set of taxonomy is very useful for the development of the MLCM taxonomy. However, due to the difference in the underlying incident causation model and research objectives, their taxonomy was adopted with many modifications.

There were also several works in the human error areas (Reason 1990; Rasmussen 1982) where the classification requires cognitive information that is frequently missing or inconclusive in construction accident investigation reports. These classifications often require expertise and resources that are not readily obtainable in the construction industry. However, human error classifications that focus on behavioural aspects, for example the taxonomy developed by Swain and Guttman (1983), can be more easily adapted into the substandard act component of the MLCM.

A draft set of taxonomy based on the taxonomy used by OSD and the literature review was first used to analyse forty accident cases. Following that, the taxonomy was evaluated and changes were made based on the evaluation. During the actual analysis, the taxonomy was constantly re-evaluated and minor changes were made as and when it was deemed necessary. The main categorisation of the taxonomy is summarised in Figure 3.5, and the full list of taxonomy is shown in Appendix 1.

In the MLCM taxonomy presented in Figure 3.5, the type of work that the main participants of the incident were involved in is used as the situational variable. The SMS components stated in the SMS failures section is based on the SMS elements described in Singapore’s code of practice for SMS for construction worksites (CP79) (PSB 1999). If necessary the section on SMS failures can be replaced by the main elements of any organisation’s SMS structure.

The results of the study on the 140 fatal accident cases are depicted in the eight histograms (Figures A2.1 to A2.8) in the Appendix 2. The eight histograms show the distribution of the type of work (situational variable), contact event, breakdown event, substandard acts, substandard physical conditions, immediate personal factors, job factors based on job functions and job factors related to site management.

Figure 3.6 shows a summary of the results depicting the main contributors in each component of the MLCM. In terms of type of work that resulted in fatal accidents, structural work and architectural/renovation/finishing work made up 57.9%. The results might be affected by the more frequent occurrence of the two types of work. Still, the figure warrants greater attention to be given to both types of work.

1. Situational Variables- Type of Work			
1.1	Architectural/Renovation/Finishing work	1.2	Building services work
1.3	Geotechnical work	1.4	Material/equipment handling/transportation
1.5	Plant/ machinery/ equipment maintenance/ dismantling /installation	1.6	Structural work
1.7	Other types of work		
2. Types of Contact Event			
2.1	Fall of person- Person struck object	2.2	Struck by falling objects
2.3	Striking against or struck by objects	2.4	Caught in or between objects
2.5	Over-exertion or strenuous movements	2.6	Exposure/contact with extreme temp/pressure
2.7	Exposure/contact with electric current	2.8	Exposed to harmful substances/radiations
2.9	Other types of incidents		
3. Types of Breakdown Event			
3.1	Collapse/toppling of object	3.2	Lost of balance- Fall of person
3.3	Object fall off surface	3.4	Loss control of plant/vehicle (Runaway plant/vehicle)
3.5	Collision between objects	3.6	Failure of equipment (breakage)
3.7	Fire/explosion	3.8	Other types of breakdown event
4. Types of Substandard Physical Conditions (Immediate Causes)			
4.1	Substandard plant/machinery/equipment/tools	4.2	Substandard construction material
4.3	Substandard structures/parts of structure	4.4	Substandard work environment
4.5	Other substandard physical condition		
5. Types of Substandard Acts (Immediate Causes)			
5.1	Extraneous Acts	5.2	Improper equipment usage
5.3	Inappropriate response to emergency	5.4	Omission of basic safety measures
5.5	Spatial error	5.6	Improper work procedure
5.7	Other substandard acts		
6. Types of Personal Factors (Immediate Causes and Underlying Factors)			
6.1	Lack of knowledge/skill	6.2	Mental/psychological factors
6.3	Improper motivation	6.4	Physical/physiological factors
6.5	Other personal factors		
7. Types of SMS Failures (Refer to CP 79 for detailed clauses) (A) Lack of measure (B) Inadequate measure (C) Inadequate execution			
7.1	Safety policy	7.2	Safe work practices
7.3	Safety training	7.4	Group meetings
7.5	Incident investigation and analysis	7.6	In-house safety rules and regulations
7.7	Safety promotion	7.8	Evaluation, selection and control of sub-contractors
7.9	Safety inspections	7.10	Maintenance regime for all machinery and equipment
7.11	Hazard analysis	7.12	The control of movements & use of haz. subst. & chem.
7.13	Emergency preparedness	7.14	Occupational health program
8. Types of Job Factors (Underlying Factors)			
8.1	Factors related to designers	8.2	Factors related to operatives
8.3	Factors related to project management/corporate	8.4	Factors related to site management
8.5	Other job factors		
9. Types of Organisational Factors (Underlying Factors)			
9.1	Poor safety and/ or organisational culture	9.2	Inappropriate organisational structure
9.3	Lack of organisational learning	9.4	Lack of stable workforce
9.5	Lack of formal and informal communication structure	9.6	Other organisational factors

Figure 3.5 Main headings of the MLCM taxonomy

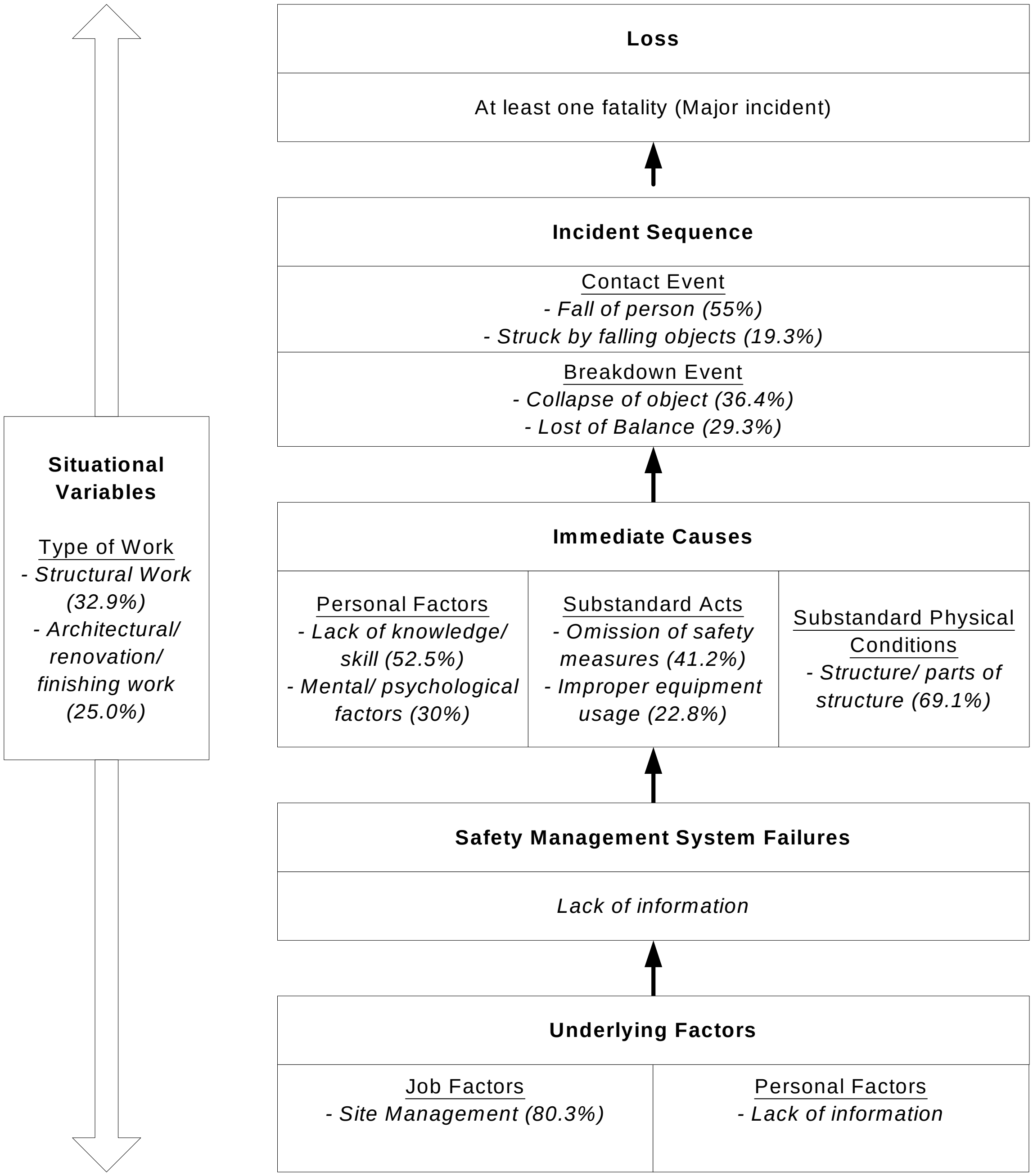


Figure 3.6 Summary of findings from 140 fatal accident cases

With respect to incident sequence, the distribution of the type of contact event is in line with results presented in several other studies based in different countries (Hinze, et al., 1998; Jeong, 1998; Kartam and Bouz, 1998). Referring to Figure 3.6, fall of person is the main type of contact event in construction industry (55%). Struck by falling object is the next highest occurring type of incident, but at a relatively lower frequency of 19.3%. As for breakdown event, collapse of object (36.4%) is the main type of breakdown event with the next highest occurrence being lost of balance (29.3%). Intuitively, the findings makes sense, as the high occurrence of lost of balance and collapse of objects naturally leads to a high occurrence of fall of persons and struck by falling objects.

Findings on substandard acts reveal a high percentage of omission of basic safety measures (41.2%) like the wearing of personal protective equipment (PPE) and checking of the vehicle's rear before reversing. The other main substandard act is improper equipment usage (22.8%); some common examples are workers using defective mobile scaffolds for work, and using employee lifts to transport construction materials. With respect to substandard physical conditions, the main violation is in substandard structure/parts of structure (69.1%). This usually refers to lack of safety structures like guardrails or barriers for open sides of buildings and shoring for trenches.

Relating to immediate personal factors, the main causes are lack of knowledge or skills (52.5%) and mental/psychological factors (30%). These personal factors are usually related to the substandard acts of operatives and workers instead of other job categories. This would indicate that training and education of operatives can be a vital link in reducing substandard acts and physical conditions. However, deeper analysis of the factors would be needed to identify the appropriate strategies.

From the analysis it is identified that there is insufficient information on the types of SMS failures. The accident investigators had not explicitly defined which part of the SMS is related to the incident. This shows a lack of effort in focusing on the SMS during the investigations.

For underlying factors, the investigators only identified the job factors. No information on personal factors that influenced those contributory underlying job factors was available. Most of the job factors belong to the category of site management (80.3%). The high concentration of site management factors shows that site management plays an important role in construction safety. A more detailed analysis is depicted in A2.8 of Appendix 2. The top three factors with the highest occurrence are: failure to ensure proper work practices/monitor site work (17.3%), inadequate inspection (16.3%) and failure to obtain/allocate adequate/proper physical resources (14.3%). The results also imply that there are some procedures and safety measures in place, but there is a lack of enforcement and proper execution. This shows that when there is a lack of close supervision on site and inadequate provision of physical resources to operatives (workers, technicians and plant operators), the SMS can fail, resulting in substandard acts by operatives and the occurrence of substandard physical conditions.

It can be seen that the accident/incident investigation focuses primarily on the identification of incident sequence and immediate factors. The lack of identification of specific SMS components that failed, represents a missed opportunity to improve the SMS so that the recurrence of similar accidents can be alleviated. Similarly, underlying personal and organisational factors were also not identified, and this lack of understanding of underlying factors can lead to ineffective safety strategies being adopted, which is detrimental to site safety performance.

The analysis has shown that the MLCM and its taxonomy could be used to systematically codify incident cases to reveal useful information and statistics that could be utilised to facilitate safety planning. More specifically, the codified incident information and statistics helps the safety planning team to: (1) identify hazards, (possible hazardous types of work and incident sequences), (2) assess the risk (likelihood and severity) of the hazards based on statistical frequency, and (3) understand the immediate causes, SMS failures, and underlying causes so as to design and plan appropriate safety measures to reduce or eliminate the risks.

3.5 Application in Safety Planning

The proposed safety planning process is also structured based on the MLCM so that the knowledge stored in incident cases can be retrieved and reused. To demonstrate how the MLCM can facilitate the second level of feedback, i.e. utilisation of past safety knowledge in safety planning (see Figure 1.2), a safety planning process for a hypothetical case is illustrated below.

3.5.1 Risk Assessment

Figure 3.7 shows a risk assessment of a lifting operation based on the MLCM framework. The situational variables are deliberately made to be similar to the incident case study presented earlier (Figure 3.4), which is highly probable as the situation is relatively common. Based on the situational variables, a risk assessment tree, structured similarly to the event tree analysis methodology, can be developed. The risk assessment tree is comprised of the possible incident sequences and their possible consequences. The incident sequences can be inserted based on past incident cases or safety plans. For instance, the shaded boxes in Figure 3.7 show the incident sequence and consequences of the previously discussed incident being inserted into the risk assessment tree. The process of forming the various branches of the risk assessment tree is in fact the SKMS hazard identification process, which will be elaborated in chapter 4.

The assignment of probabilities of occurrence for each event ($P(B_1)$, $P(C_{12})$ and $P(CSQ_{123})$ for the breakdown event, contact event and consequence respectively) can be based on subjective sources such as expert judgment or objective observations of actual incidents. The use of objective observations will require a large amount of data and an effective retrieval and adaptation system, which are rarely available currently. On the other hand, a purely subjective assignment of probability will reduce the credibility of risk assessment. The objective in the proposed MLCM-based approach is to integrate the observed incidents into the subjective probabilities assigned. The Bayesian approach (Ang and Tang 1975) can be used to provide this integration so that the prior subjective probabilities can be revised with observed occurrence of the events from incident investigations. A detailed discussion of the application of Bayesian statistics and SKMS risk evaluation will be covered in chapter 5.

This feedback of incident investigation information is facilitated by the common structure in both incident investigation and safety planning, through the consistent use of the MLCM. As a result, an augmenting tree can be developed based on past investigation cases to supplement the risk assessment tree. Figure 3.8 shows an example of such an augmenting tree developed for painting work based on the 140

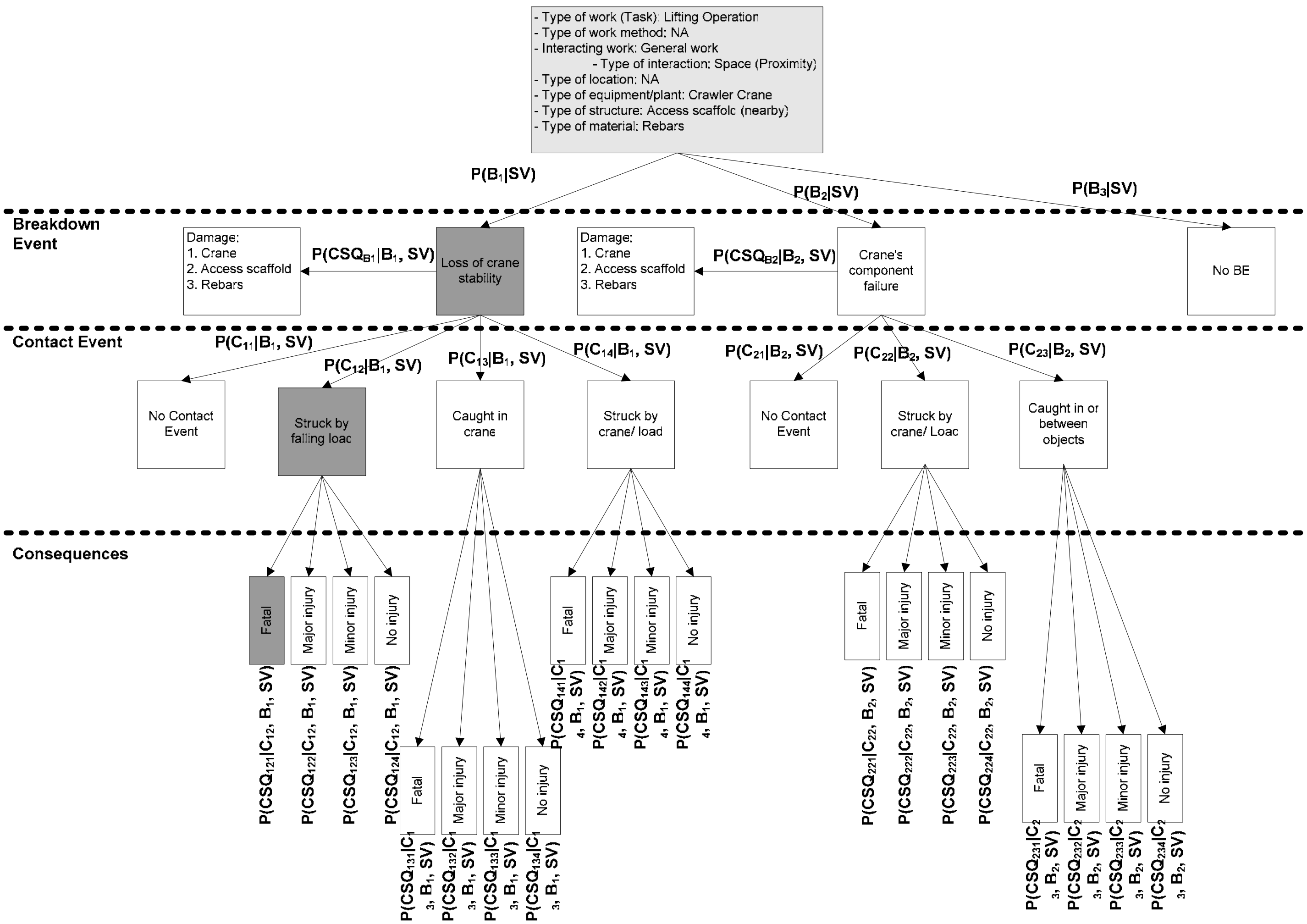


Figure 3.7 Risk assessment based on MLCM

investigation cases. The figure summarizes the incident sequences of past painting work incidents and their corresponding number of occurrences. For instance, out of the ten accident cases related to painting work, five cases have a breakdown event of the "Lost balance" category, two cases of the "Loss control of transport/plant" category, and the remaining three of the "Collapse of temporary structure" category. Further classification could also be made to provide more details on the category of the breakdown event. For example, out of the three cases with breakdown event of the "Collapse of temporary structure" category, two involved a lifting platform and one involved a mobile access scaffold. Similarly, contact event and consequences for each of the incidents can also be classified based on the set of MLCM taxonomy, and the number of occurrences in each category can be determined as depicted in Figure 3.8. During risk assessment, the augmenting tree serves to facilitate the identification of a possible incident sequence (breakdown event, contact event and consequences), and ensures that past incident occurrences will not be overlooked. Furthermore, the assignment of probability of occurrence can also be guided by relative frequencies derived from the numbers denoted in the augmenting tree, thus providing a more objective basis for probability figures.

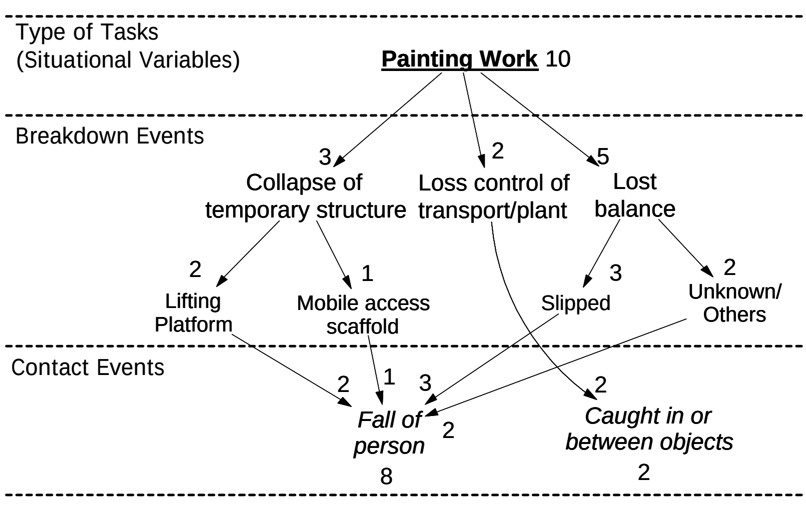


Figure 3.8 Augmenting tree developed based on 140 accident cases

Similarly, past risk assessment trees developed in past safety plans can also be reused in the same way augmenting trees are used. In this way, incident sequences that had never occurred before, but were identified by previous safety planning teams, will also be included. Furthermore, the probabilities assigned in a past risk assessment tree can be used to provide the prior estimates of probability or frequency values of incident events. These prior estimates can then be updated based on available objective data using the Bayesian approach as mentioned earlier.

3.5.2 Risk Control Selection

Figure 3.9 shows the risk control selection at the preventive level for a possible breakdown event, “Loss of crane stability”, during the crane-lifting operation presented in the earlier section on risk assessment (see Figure 3.7). In order to select the relevant risk controls, the immediate causes of the breakdown event are first identified based on an approach similar to the preceding risk assessment. Due to the similarity of the situational variables and breakdown event between the past case of Figure 3.4 and the new case described in Figures 3.7 and 3.9, the findings from the investigation of the past case may be utilised. In particular, the immediate causes identified, i.e. the substandard act “Overloading” due to the personal factor, “Underestimation of load” is included as a possible immediate cause of the breakdown event “loss of crane stability” in the new case.

Furthermore, the preventive measures recommended by investigators are also retrieved. In this example, the retrieved preventive measures include procedures to ensure that weight of load is communicated prior to lifting, and also the appointment of a lifting supervisor (general preventive measure). In this way, preventive measures

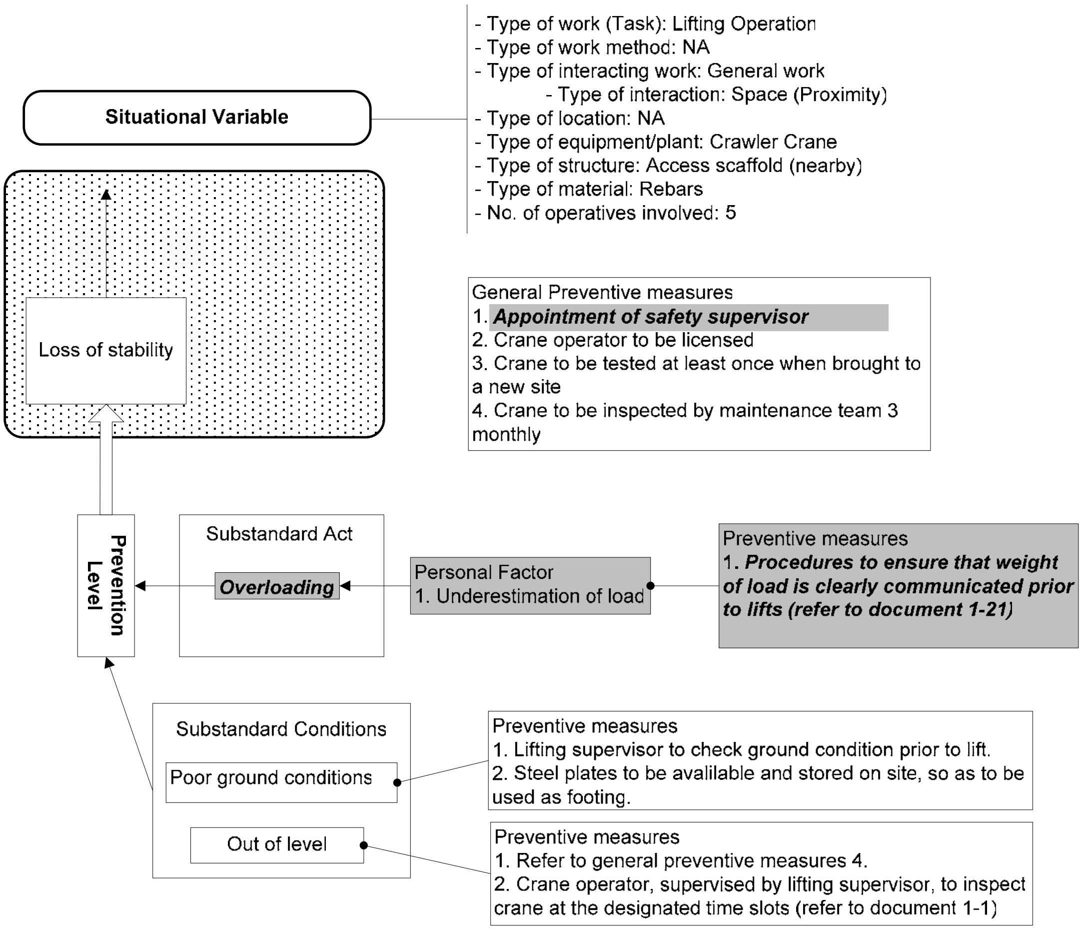


Figure 3.9 Risk control selection based on MLCM

need not be developed from scratch; instead if possible, they could be retrieved from past incident investigations when the situational variables for the new case resemble those of past cases. These measures would usually be practical and effective since they have been implemented before. Moreover, the retrieved measures can be adapted to better accommodate any unique situation of the present case, and also improved on, if necessary. It should be noted that the safety planning team need not be constrained by past cases and they can also identify immediate causes, and corresponding measures based on their knowledge and experience. For instance, the substandard condition, "Poor ground conditions", and corresponding control measures such as checking of ground conditions by lifting supervision and use of steel plates as footing, (Figure 3.9) could have been identified and planned based on pre-emptive evaluation of the situation.

The above discussed approach can also be adopted for the other levels of intervention, i.e. pre-contact and post-contact levels. In this way a thorough and systematic SMS can be developed.

As in the case of risk assessment, past safety plans can also be retrieved and reused in the same way past incident cases are reused. The reuse of past safety plans will allow safety knowledge of other safety planning teams to be utilised and possibly filling in gaps that occurs due to the lack of incident cases in certain areas. Furthermore, the example only shows an incident case being retrieved. In an actual situation, more than one incident case together with past safety plans can be included or addressed. As the organisation accumulates more safety plans and incident cases, the number of retrieved and utilised cases will increase. Through the organisational learning process of "remembering" and applying past safety knowledge, the quality and efficiency of the safety planning process will improve over time.

3.6 Conclusions

This chapter presented an incident causation model, the Modified Loss Causation Model (MLCM), which is meant to facilitate feedback at two levels, firstly, to the SMS that had failed, and secondly, to the safety planning process for future construction projects. Through the two levels of feedback, construction SMS, and hence safety performance of the industry, can be continually improved.

In order to achieve the two levels of feedback, the MLCM is designed to provide a systematic and logical structure for both incident investigation and safety planning, such that if the MLCM is applied consistently, the depth and breadth of both the processes will be ensured. Through the use of the MLCM as a common model for incident investigation and safety planning, incident investigation information and past safety plans can be retrieved and utilised in new safety plans. To facilitate the storage and feedback of safety knowledge structured based on the MLCM, a set of taxonomy was also developed. The taxonomy was successfully applied on 140 fatal accident cases and it will be used to guide the development of the knowledge representation scheme in chapter 4.

However, in order to fully exploit the ideas and concepts illustrated in this chapter, a computer-based system will have to be developed. Thus the subsequent chapters will present the various components of the SKMS which will facilitate the risk assessment process detailed in Section 3.5.1. The SKMS components presented in this thesis will also form the basis for the implementation of risk control selection process discussed in Section 3.5.2.

Chapter 4

KNOWLEDGE REPRESENTATION AND CASE RETRIEVAL

4.1 Introduction

The Modified Loss Causation Model (MLCM) is meant to act as a common knowledge framework for the SKMS. However, for the SKMS to be implemented based on Case-based reasoning (CBR) concepts, an integrated and detailed knowledge representation scheme has to be developed. This knowledge representation scheme is very important because it forms the basis for both case retrieval and adaptation (chapter 5).

This chapter is organised into two main portions. The first portion presents the knowledge representation scheme employed to abstract knowledge in incident cases and past risk assessment trees. The knowledge representation scheme is implemented in a relational database and the design of the database will be presented. The second portion details the case retrieval mechanism and how it is implemented in the SKMS. The discussion on case retrieval will be focused on the similarity scoring functions.

4.2 Knowledge Representation of Incident Cases and Risk Assessment Trees

In Case-based Reasoning (CBR), a case usually contains two broad types of knowledge: (1) the lessons that it teaches, and (2) the context in which it can teach those lessons (Kolodner 1993). In this research, the “lessons” that each case teaches consists of the incident sequences and the risk (probability × severity) posed by them, whereas the context is described by the situational variables (refer to Figure 3.1). Both types of knowledge have to be represented carefully to ensure appropriate retrieval, adaptation and application of past cases.

4.2.1 Modelling Approach for the Lessons Learned

For a number of CBR concepts that handle real world problems, cases are made up of several inter-dependent sub-lessons or sub-cases. Kolodner (1993) identified two possible approaches in modelling and utilising these sub-cases. The first approach splits and stores a case or episode into independent smaller cases that are sub-parts, or snippets, of the original case (Kolodner 1988). Each snippet is a set of specific knowledge with its own indices so that these snippets can be retrieved separately from the main case that it belongs to. The snippets are then pieced together to form a relevant "lesson" for the user. Links between snippets and the overall case may need to be preserved to maintain the structure of the reasoning and also the completeness of each full episode. The second approach employs monolithic cases which are essentially "large" cases that keep all the snippets intact during retrieval. The "large" case's indices can be clearly associated to different snippets of the case so as to allow removal of irrelevant snippets after retrieval.

Even though the two approaches are different, both agree that each "large" case can be separated into snippets. In the context of the SKMS, a "large" case would naturally refer to a risk assessment tree or an incident case. Within each "large" case, incident events (breakdown event, contact event or consequence) would be a suitable snippet, because they form the most basic knowledge blocks of any incident sequence. An incident case would contain only one incident sequence, i.e. one breakdown event, one contact event and their consequences, while a risk assessment tree would contain a set of incident sequences that forms a tree structure (see Figure 3.7). However, both incident cases and the risk assessment tree can be easily represented by incident events.

The main difference between the two approaches occurs during retrieval. For the monolithic case approach, the risk assessment tree or incident cases are retrieved and then irrelevant incident events are removed or adapted to improve relevance. On the other hand, the snippet approach retrieves only relevant incident events and uses these incident events for risk assessment. Past researches (for example Redmond 1990a, 1992, Kolodner and Simpson 1989) have shown that both representation approaches are feasible.

The monolithic approach was implemented in the SKMS because it was recognised that case representation is at best an abstraction of a real-world episode, and it would be more prudent to keep a "large case" intact and not separated into sub-cases or snippets. In this way, subtle details within a complete case which could have been missed if the case has been separated into snippets will be made available to the human user. Furthermore, the monolithic approach also reduces the computational cost of retrieval, because the number of cases to be searched and assessed increases tremendously when each snippet is treated as an individual case. Besides, it is also more natural for risk assessment teams and incident investigators to view each incident or risk assessment as an episode or scenario. However, it is noted that both approaches are viable and the snippets approach could still be implemented.

4.2.2 Modelling Approach for the Context of Lessons Learned

The modelling of the context in which the lessons are applicable in is also known as the indexing problem. The indexing problem can be tackled at two levels, firstly, selection of an appropriate indexing vocabulary, and secondly, the selection of specific indices for each case (Kolodner 1993). Indexing vocabulary is a set of possible descriptors that can be used to index all the cases in the case base, while indices are specific descriptors that designate the situations under which the case is relevant. In this research the indexing vocabulary will correspond to the situational variables of the MLCM (see Figure 3.1).

4.2.2.1 Indexing Vocabulary

In the SKMS, the risk assessment process is based on the Job Hazard Analysis (JHA) (also known as Job Safety Analysis) approach. Thus, the indexing vocabulary is designed to be able to support the JHA's structure.

A JHA is a common safety planning technique that is focused on a specific job and the analysis begins by separating the job into specific job steps. For example, a lifting operation of precast element (the job) can be separated into five basic job steps: (1) position lifting gear over precast element, (2) rig-up precast element, (3) lifting and positioning of precast element, (4) un-rig precast element, and (5) lifting of lifting gear away from precast element. Each job step is then evaluated for the possible hazards and their risks. Subsequently, relevant risk controls are then selected to eliminate or reduce the risk. Since incidents usually occur during a specific job step at some stage in an activity, an indexing vocabulary that can describe the situational variables of a job step during JHA can also be used to describe the context of an incident.

In this research, the indexing vocabulary employed is based on the following linguistic structure:

Action(s) executed on *object(s)-worked-on* using *resource(s)* at *location(s)* with *nearby object(s)* and *nearby action(s)*

For example, when the indexing vocabulary is applied on the example given in Section 3.4.2 (chapter 3), the situational variables would be: *Lifting operation (Action)* executed on *rebars (object-worked-on)* using *crawler crane (resource)* at *unspecified (location)* with *nearby access scaffold (nearby object)* and *general work (nearby action)*.

Each of the underlined italics terms in the above linguistic structure are in fact potential sources of harm that can contribute to the occurrence of incident sequences in a job step. As highlighted in chapter 3, "harm" is usually due to an uncontrolled source of energy or substance being released. Logically, this source of energy or substance would originate from: (1) human actions that are applied during some course of work ("action" or "nearby action"), (2) any object or substance that was used to facilitate work, was acted upon or spatially close to the human action ("object-worked-on", "resource" or "nearby object"), or (3) the environment or location in which the job was being executed in ("location"). Thus, the indexing vocabulary used in this research, "action", "object-worked-on", "resource", "location", "nearby object", and "nearby action", also corresponds to hazards that can lead to occurrence of incident sequences.

4.2.2.2 Indices

During the indexing of a particular case, not all six types of indexing vocabulary have equal importance in relation to the lessons of the case. For instance, Figure 4.1 shows an accident case based on an actual accident report obtained from Singapore's Land Transport Authority. The accident occurred during the lifting of a precast segment that was lowered into an excavated area. Timbers were being used to act as cushions or pads to protect the precast segments. After the rigger rigged up the precast segment, the crane lifted the precast segment. However, the crane operator and rigger did not realise that a piece of the timber cushion (2” × 2” × 2 feet) was stuck onto the precast segment. As the precast segment was being lifted the timber cushion fell off (breakdown event) and struck the rigger (contact event), causing his nose to bleed and hence 0.5 man-days lost (consequences).

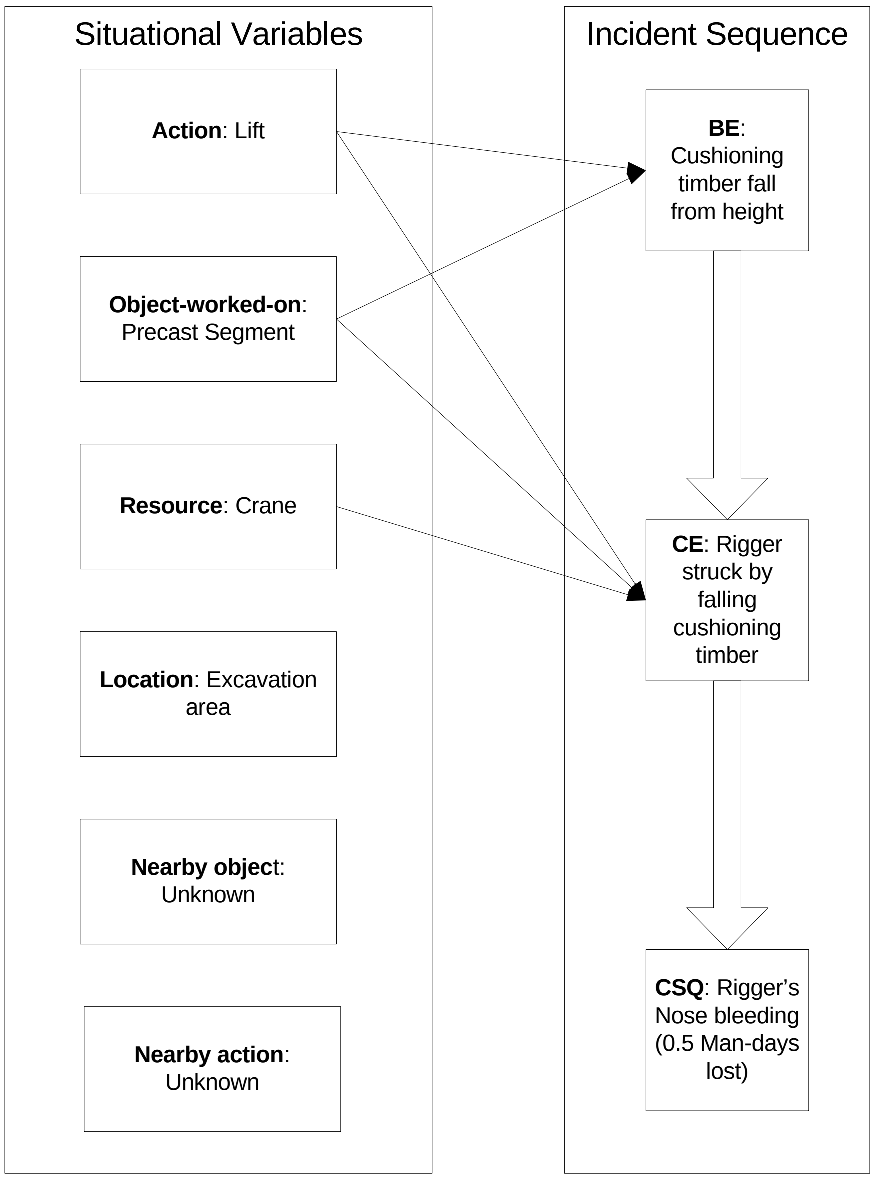


Figure 4.1 Example of indices chosen for an incident sequence

The situational variables of the case are represented in the left portion of Figure 4.1, which can also be expressed in the following job step description:

Lifting of *precast segment* using *crane* at *excavation area* (with *unknown nearby object* and *unknown nearby action*).

However, as indicated by the arrows in Figure 4.1, only the action, object-worked-on and resource are related to one or more of the incident events. In this particular case, the breakdown event, "Cushioning timber fall from height", could only have happened when the lifting action was executed and that the cushioning timber was present. Furthermore, the cushioning timber was present mainly due to the object-worked-on, i.e. the precast segment. Hence, for the breakdown event, the indices or necessary situational variables are the action (lift) and the object-worked-on (precast segment). It is noted that despite the fact that the crane is also a resource, it is not related to the breakdown event directly or indirectly. Thus, crane was not a necessary situational variable.

Similarly, the contact event, "Rigger struck by falling cushioning timber", could only have occurred because the rigger and the falling cushioning timber were present. The rigger was present because he was involved in the preparation for the lifting of the precast segment using the crane, while the necessary situational variables for the falling cushioning timber are related to those of the breakdown event. Hence, besides the action (lift) and object-worked-on (precast segment), the necessary situational variables for the contact event will also include the resource (crane).

Thus, as a whole, action (lift), object-worked-on (precast segment) and resource (crane) are important situational variables for the case. However, not all of these indices are of equal importance, for instance, in the earlier example of Figure 4.1, the resource, crane, is not a necessary situational variable for the breakdown event (BE) (therefore no arrow linking crane and the BE), and it is also not the only necessary situational variable for the contact event (CE) (action and object-worked-on are also necessary situational variables for the CE). Thus the importance of the resource to the whole case is relatively lower than the action and object-worked-on situational variables, which are necessary situational variables for both BE and CE. Other non-missing situational variables such as "Location- Excavation Area" are not necessary situational variables for any incident events and thus have low weights. Still these situational variables provide contextual information that would give a richer picture of the case. A match on these contextual situational variables would also mean higher similarity, but such similarity is of lower significance and it should not cause distortion in the similarity scoring function. During retrieval, the differences in importance of different situational variables need to be accounted for and this will be discussed in a later section.

There can be more than one value for each type of situational variable. For example, an action that uses two resources or handles two objects-worked-on simultaneously will require more than one resource or object-worked-on to be indexed. Besides, the links between the situational variables and each individual incident event (snippets) will have to be maintained, so that the relevance of a specific event can be assessed during adaptation. The specifics of adaptation processes will be covered in chapter 5.

4.2.3 Implementation of the SKMS Case Base

The case base of the SKMS is implemented in Microsoft Access 2002 (MS Access), which is a relational database management system (RDBMS) that has a programming language called Visual Basic for Applications (VBA) embedded in it. MS Access and VBA are relatively easy to use and have various functions to facilitate quick prototyping. By implementing the case base in MS Access, other components of the SKMS such as the similarity scoring functions can also be developed using VBA.

Figure 4.2 shows the core relational design for the SKMS case base. Each case's situational variables are stored in the table, "tblSitVar", using attribute-value pairs, "SitVar" and "SitVarValue". "tblEvent" represents an individual event of an incident case, i.e. it is a snippet of the case. Each event is linked to other events through the "PreEventID" field to form the risk assessment tree or incident sequence, i.e. the "large case". The same "attribute-value pair" approach is adopted for the representation of the incident sequences of cases, where "EventType" is the "attribute" and "EventValue" contains the "value". The "EventValue" is broadly based on the taxonomy described in Appendix 1, while the "EventType" can be breakdown event, intermediate event, contact event or consequence. Intermediate event is very similar to a breakdown event, but it occurs as a result of a breakdown event and prior to a contact event. It was included to allow more flexibility during modelling of incident sequences.

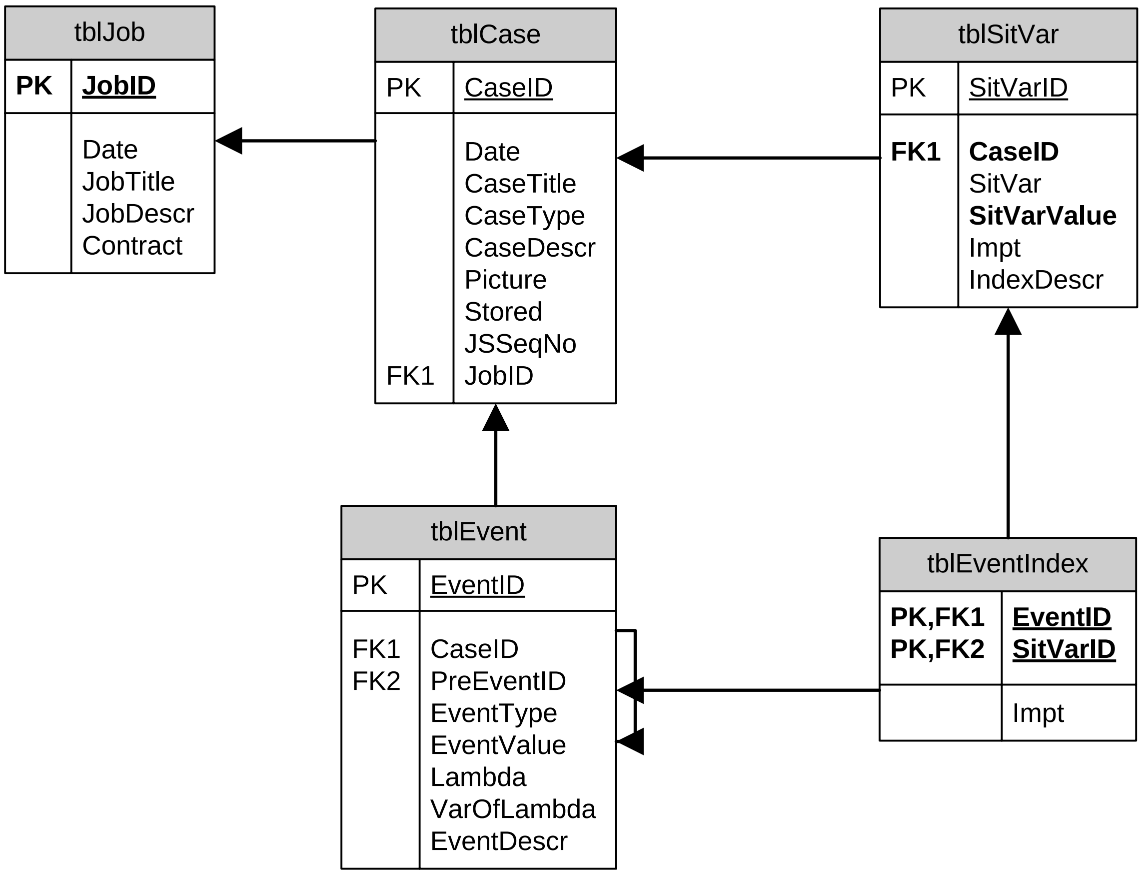


Figure 4.2 Relational Design of the SKMS Case Base

Another important field in the table "tblEvent" is the field, "Lambda". It is the estimated frequency rate of occurrence, $\bar{\lambda}$, of an event, which is the number of incidents per 50,000 man-hours worked. $\bar{\lambda}$ represents the estimated frequency or likelihood of occurrence of job steps or incident events, and it is necessary for the estimation of the risk of different incident sequences. Accompanying $\bar{\lambda}$ is the field, "VarOfLambda", which stores the variance of λ, and is also necessary for the Bayesian updating of likelihood estimates. Chapter 5 will elaborate on how $\bar{\lambda}$ and the variance are utilised during risk analysis.

As can be observed in Figure 4.2, each event is also related to specific indices of the “large case” through the link table, “tblEventIndex”. This many-to-many relationship allows adaptation to be done on the “large case” to make sure that irrelevant parts of the “large case” can be removed. The adaptation process will be presented in chapter 5.

4.3 Case Retrieval

In a CBRS, the knowledge representation scheme directly affects the quality of case retrieval and hence the overall effectiveness of the system. In the SKMS, cases are represented and retrieved using the monolithic approach. The following subsections will present how the earlier discussed knowledge representation approach is employed in the proposed case retrieval strategy.

4.3.1 Overview of Case Retrieval Approaches

Retrieval of past cases is one of the most important processes of any CBRS. The quality of retrieval directly affects the relevance of retrieved cases and hence the overall quality of the reminding capability of a CBRS. Two main types of retrieval approaches are usually employed: indexing approaches and similarity scoring (or distance-based) approaches (Liao et al. 1998).

Indexing approaches organise cases based on an indexing structure that is derived using various machine learning methods, for example decision tree, neural network and clustering algorithms. During the retrieval, the system will then traverse the indexing structure and search for the stored cases that match the input case’s indices. However, to develop the indexing structure, there is a need for a relatively large number of cases which should cover as wide a spectrum of cases as possible. In most organisations, there may not be a sufficient number and variety of cases to meet this requirement. Besides, indexing approaches generally require a clear output or outcome variable so as to allow the indexing structure to learn from the outcome of past cases. In the case of the SKMS, there is no clear outcome variable that allows a decision tree or neural network to be built. Furthermore, to ensure that the indexing structure is relevant and effective, the indexing structure may have to be retrained or redeveloped when the case base grows. This requires more effort in maintenance and may defeat some of the benefits of a CBRS. Thus, indexing approaches are not suitable to be applied in the SKMS.

CBRS are also sometimes known as "similarity-searching systems" (Liao et al. 2000), that is because most CBRS use the similarity scoring approach. Similarity scoring approaches compute a quantitative distance or similarity score between the input case and each stored case during retrieval. The similarity score is used to determine relevance of stored cases and realise inexact matching, where the higher the similarity score the more relevant the stored case. Subsequently, the top *K* number of cases will be retrieved for utilisation or further adaptation. Thus, the similarity scoring approach is also known as the *K* nearest neighbours (KNN) retrieval. The key advantage of using similarity score is the flexibility of the approach. Furthermore, the approach can even be applied on relatively small case bases.

Similarity scores are usually determined at two levels: local similarity and global similarity (Empolis Knowledge Management 2001). Local similarity refers to the similarity between the values of a particular attribute (or situational variable, in the case of the SKMS) of two cases. On the other hand, global similarity refers to the similarity between two cases. Local similarity is usually determined using a similarity function, and global similarity is determined through a weighting function that places different importance on local similarity scores of different attributes. The determination of global similarity score will be discussed subsequently in this chapter.

A wide range of similarity functions have been developed and utilised in numerous fields such as machine learning, data mining, and statistics to determine local similarity. Similarity functions can be categorised based on the type of attributes that it operates on. The categories generally include interval-scaled, ratio-scaled, binary, nominal, and ordinal attributes (Han and Kamber 2001).

Interval-scaled attributes are basically continuous measurements on a linear scale. Examples of interval-scaled attributes include such things as length in kilometres, height in metres and weight kilograms. Ratio-scaled attributes refer to attributes that are measured based on a nonlinear scale, such as an exponential scale. A binary attribute is an attribute that has only two possible values: 0 or 1. For example, the attribute "CaseType" of the table "tblCase" in Figure 4.2 can be represented as a binary attribute, where the possible values are "incident" (0) or "risk assessment tree" (1). Nominal attributes can be considered as an extension of binary attribute, where the key difference is that a nominal attribute can have more than two possible values or state. For example, the situational variable, resource, can have a very large number of possible values, ranging from crane, to pneumatic breaker to arc-welding set. It is noted that all the indices in the SKMS that are used to determine similarity scores are nominal attributes. These attributes are the various situational variables that are represented using the attribute-value pair, "SitVar"-"SitVarValue" (Figure 4.2). Lastly, ordinal attributes or more specifically discrete ordinal attributes, are nominal attributes that are ordered in a meaningful sequence. An example of an ordinal attribute would be size, which can have the possible values, small, medium and large.

4.3.2 Similarity Functions for Nominal Attributes

Most nominal attributes employ taxonomy-based approaches for similarity assessment, where a taxonomy tree (Empolis Knowledge Management 2001; Cognitive Systems 1992) or an abstraction hierarchy (Kolodner 1993) is the basis of the similarity score. Figure 4.3 shows an example of a taxonomy tree that is developed based on Illingworth's (2001) categorisation of construction plants. For taxonomy-based approaches, the degree of similarity between two values is determined based either on the distance of the most specific common abstraction (MSCA) of the two values or the specificity of the MSCA. The MSCA can be defined as the most specific or lowest level value on a taxonomy tree that is linked to the two values that are assessed for their similarity. For example, with reference to Figure 4.3, the MSCA for "wheel barrows" and "concrete pumps" is "on-site material handling". Generally, the nearer or more specific the MSCA, the more similar the two values are.

For the specificity approach, local similarity score can be assigned arbitrarily based on how specific a MSCA is. For instance, with reference to Figure 4.3, all plants that share the MSCA, "Three-dimensional horizontal plane", can be assigned a local similarity score of 0.6, and all plants that share the MSCA, "Crane", can be assigned a local similarity score of 0.8 to reflect the fact that "Crane" is a more specific abstraction than "Three-dimensional horizontal plane".

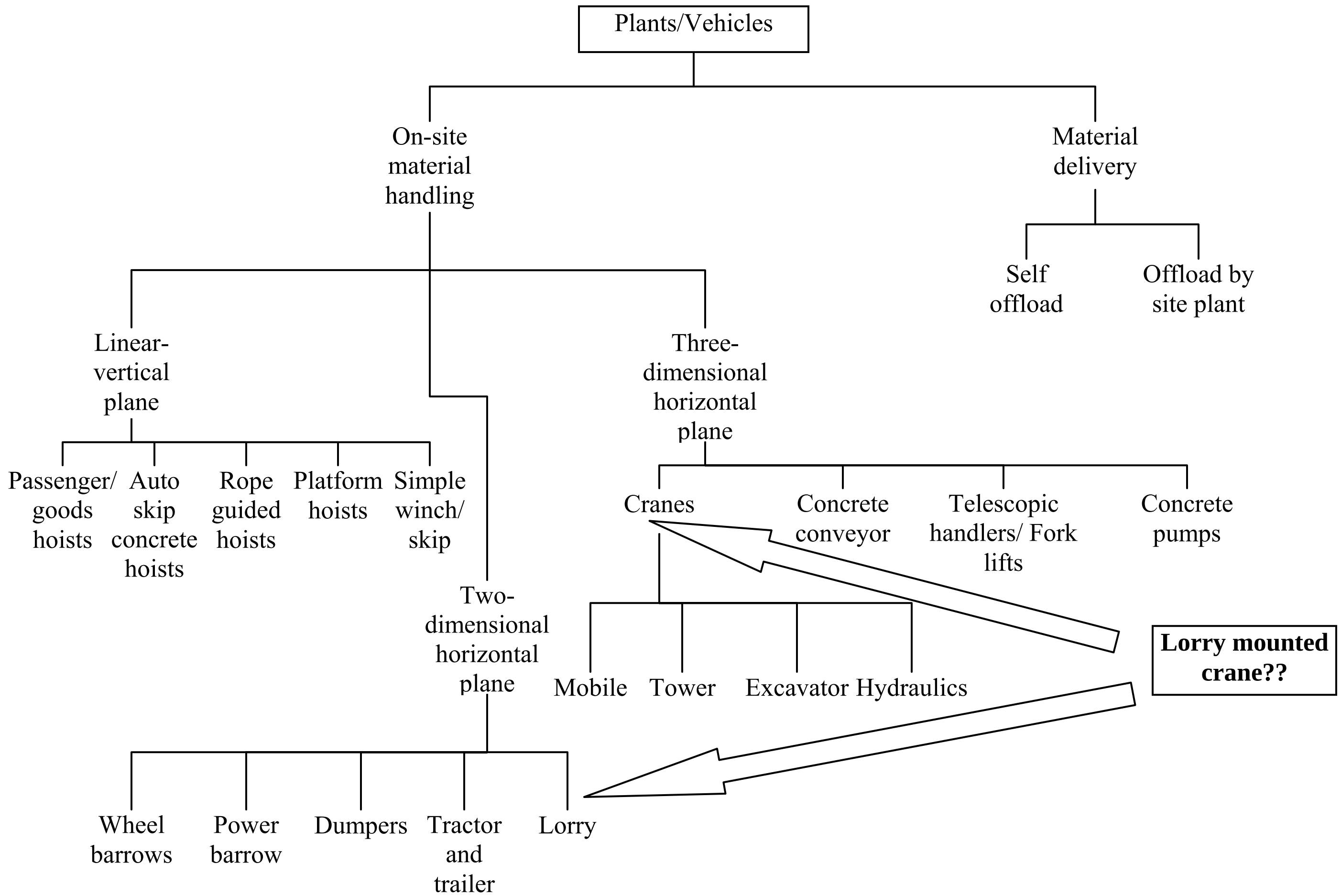


Figure 4.3 A taxonomy tree for construction plants (Illingworth 2001)

In contrast, the distance approach is usually based on Equation (4.1),

$$LSS\ (V_1, V_2) = 1 - \{1 \times [\ (L_1 + L_2) / (L_{T1} + L_{T2})\ ]\ \} \tag{4.1}$$

where LSS (V_1, V_2) is the local similarity score between values V_1 and V_2, L_1 (L_2) is the number of links between the MSCA and V_1 (V_2), and L_{T1} (L_{T2}) is the total number of links between V_1 (V_2) and the top node.

Essentially, Equation (4.1) determines the LSS between two values based on the number of links or distance from the MSCA in relation to the distance of the values from the top node. For example, with reference to Figure 4.3, the LSS between an excavator (V_1) and platform hoist (V_2) would be 0.29 (MSCA = “On-site material handling”, $L_1 = 3$, $L_2 = 2$, $L_{T1} = 4$ and , $L_{T2} = 3$), while an excavator (V_1) and a mobile crane (V_2) will have a similarity of 0.75 (MSCA = “Cranes”, $L_1 = 1$, $L_2 = 1$, $L_{T1} = 4$ and , $L_{T2} = 4$).

Both distance approach and specificity approach are not able to handle situations where a value can be classified in more than one way. This is due to the nature of taxonomy trees, which usually have strict classification rule for each value, i.e. each value can only be linked to one higher node. The problem is illustrated in Figure 4.3, a lorry mounted crane can be considered both as a crane and a lorry, and therefore it can be classified both under crane or lorry. If the lorry mounted crane is classified under both crane and lorry, there will be more than one possible LSS when the lorry crane is compared to another value. This would then cause ambiguity as to which LSS should be adopted. A taxonomy-based approach, therefore, is suitable only when the possible values can be uniquely classified, but this is often not the case in most real world problems where values usually have more than one way of classification.

In several CBR applications, nominal attributes are also transformed into binary attributes by creating a new binary attribute for each of the possible nominal values (Yau and Yang 1998; Liao et al. 2000; Han and Kamber 2001). For example, Yau and Yang (1998) transformed the attribute "Soil Strength" into eleven binary attributes, such as "Soil_Strength_Soft" and "Soil_Strength_M_Firm", each with possible values of "0" and "1". Another alternative similarity assessment approach commonly adopted is the use of tables or rules to directly assign similarity scores between each possible pair of values (Karim and Adeli, 2003; Luu et al. 2003). However, such approaches are only feasible if the possible values of each attribute are minimal, and do not require a large number of binary attributes or unwieldy tables or rules.

Another possible approach is to evaluate the sub-attributes or sub-concepts of each of the values (Domeshek 1991b; Kolodner 1993). Kolodner (1993) gave an example of how different dishes can be compared based on the ingredients (sub-concepts) that they share and the ingredients that they differ. The local similarity score can then computed using the following equation (Liao et al., 1998),

$$\text{LSS}\ (V_1, V_2) = \alpha \times common \,/\, (\alpha \times common + \beta \times different) \qquad \textbf{(4.2)}$$

where *common* (*different*) represents the number of sub-concepts that are shared (not shared) between V_1 and V_2, and α and β are corresponding weights for *common* and *different*. It is noted that "*common*" can also be termed as intercept (∩) and "*common* + *different*" can also be interpreted as union (∪) (Cognitive Systems 1992).

The sub-concepts approach is flexible and does not require taxonomy trees to be developed explicitly. Values do not need to be strictly defined under any category, but rather each value is represented by a list of sub-concepts. However, Equation (4.2) does not take into account the difference between specific and general sub-concepts. This is detrimental to the accuracy of the similarity score because a match on a more specific sub-attribute should indicate a higher similarity, while a match on a more general sub-attribute should carry less significance.

4.3.3 Similarity Scoring in the SKMS

Based on the above discussion, it is apparent that none of the local similarity assessment functions or approaches are directly germane to the context of the SKMS. Thus, a modified approach that is based on the sub-concept approach has been developed. For the proposed approach, a semantic network (Russel and Norvig 1995), instead of a taxonomy tree, is constructed for various situational variables. An example of a semantic network for the situational variable "Action" is shown in Figure 4.4. A semantic network is an extension of a taxonomy tree, where the key difference is that nodes in a semantic network can have more than one parent. For example, the node "(2) Move" of Figure 4.4 has two parent nodes, "(1-1) Human" and (1-2-5) Self-propelled object". The semantic network allows for more flexibility during classification of concepts or sub-concepts and would not have the problem of a taxonomy tree where each value can only have one parent node.

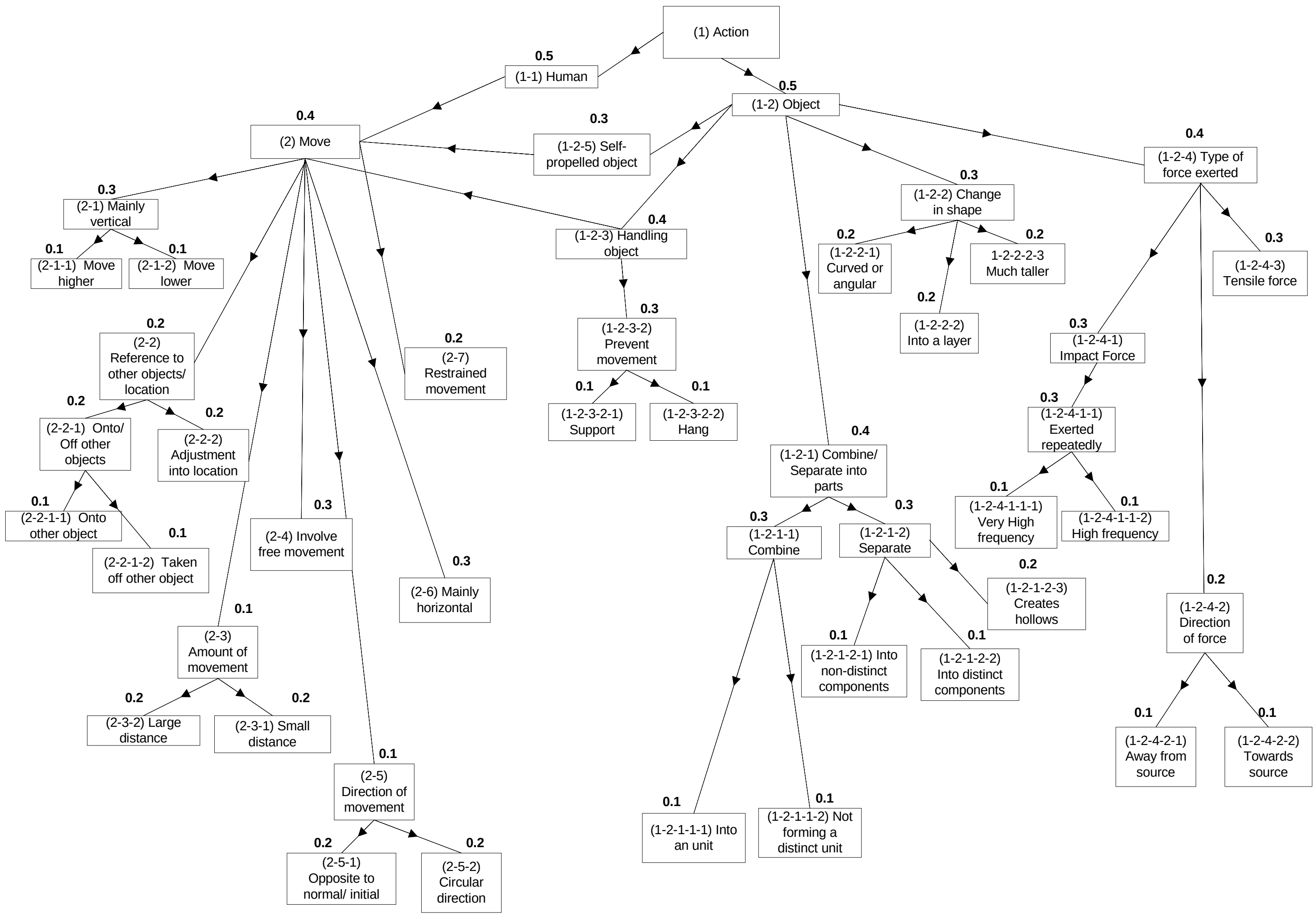


Figure 4.4 Semantic network for situational variable "Action"

The semantic networks are constructed based on the procedure depicted in the flowchart of Figure 4.5. A case base is first constructed based on more than 100 incident cases taken from the Safety Information System of Singapore's Land Transport Authority (LTA) and 10 risk assessment trees obtained from the Mine Safety and Health Administration (MSHA 2004), LTA and various construction contractors. The incident and risk assessment tree case base is then queried and all the possible values for each situational variable are collated. Each of the values is then given a working definition that highlights the key characteristics of the value that distinguishes it from the other possible values, and at the same time emphasises the usual types of hazardous object, energy or harmful substance associated with the value. For example, the value, "Lift" of the situation variable "Action" is defined as, "to move an object from a lower to a higher position, hence accumulating gravitational potential energy and producing kinetic energy during the movement." This definition highlights the nature of the action, and at the same time makes reference to the type of energy or harmful substance that is produced or accumulated during the action.

Based on the working definitions, the values are then compared and contrasted to identify similar and contrasting sub-concepts that the values represent. These sub-concepts are then repeatedly evaluated to identify more specific concepts. Related sub-concepts are then linked together by an arrow, where the arrow points from the more general sub-concept to the more specific sub-concept. As mentioned earlier, in a semantic network a lower level sub-concept can be linked to more than one parent. However, such multiple linkages can cause the semantic network to be complex and incomprehensible, and is avoided whenever possible.

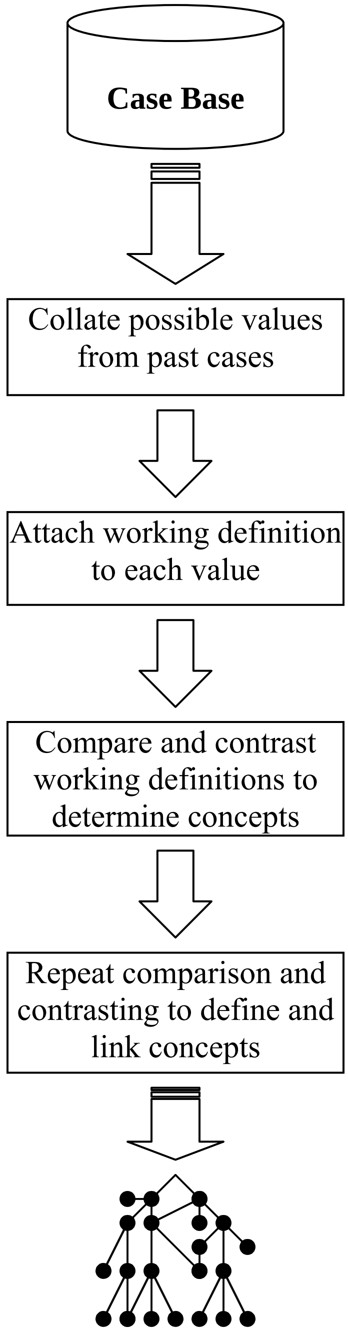


Figure 4.5 Flowchart for construction of taxonomy tree

The semantic networks developed represent all the possible sub-concepts for each situational variable. Each value can then be represented by a list of sub-concepts picked off from the corresponding semantic networks. For example, Figure 4.6 shows the list of sub-concepts for the values “Hack”, “Excavate” and “Extract” of the situational variable “Action”.

In order to differentiate the sub-concepts in terms of specificity and importance, a weight is assigned to each of the nodes (see Figure 4.4). The weights assigned are not the same as the similarity values that are assigned to the MSCA or nodes in a taxonomy tree. Instead, they range from 0.1 to 0.5, and are assigned based on the guiding principle that nodes that are higher, or closer to the root, are more influential on the categorisation of values. These higher nodes are generally given higher weights. Another principle is that sub-concepts that are more directly related to potential hazards are also given higher weights. The weights for the sub-concepts of the situational variable “Action” is indicated by the bold decimals above the corresponding nodes (see Figure 4.4). Thus the weights can be viewed as the incremental similarity due to a match on the sub-concept represented by the node.

It is noted that situational variables that refer to the same fundamental subjects share the same semantic network. For example, objects-worked-on, resources and nearby objects share the same semantic network, because all three situational variables refer to physical construction-related objects. For the prototype SKMS, three semantic networks had been developed besides the semantic network for the situational variable “Action” (Figure 4.4). The other two semantic networks are depicted in Appendix 3. It is noted that

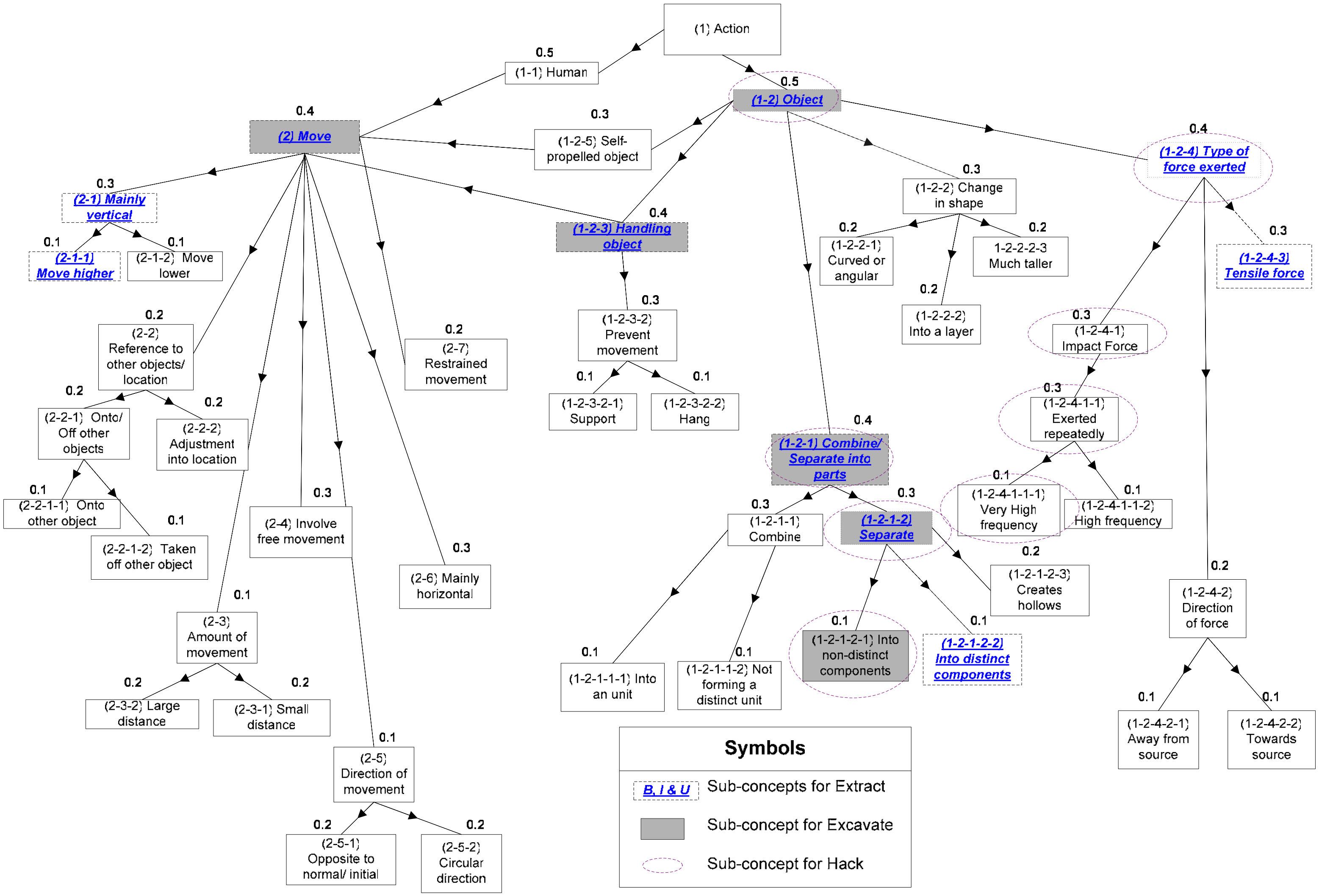


Figure 4.6 Sub-concepts for the values "Hack", "Extract" and "Excavate" under situational variable "Action"

the semantic network for the situational variable "location" is very small and forms only a taxonomy tree (see Figure A3.2 of Appendix 3). This is due to the low number of cases with the situational variable "location". When more cases is added into the case base the semantic network for "location" will also grow correspondingly.

With the semantic networks, the local similarity scores between values V_1 and V_2 can then be calculated based on the following equation,

$$\text{LSS}\ (V_1, V_2) = \sum w_{ci} \,/\, (\sum w_{ci} + \sum w_{dj}) \qquad \textbf{(4.3)}$$

where i = 1, 2, … *common,* j = 1, 2, … *different*, w_{ci} is the weight of the common sub-concept i, and w_{dj} is the weight of the sub-concept j that belongs only to either V_1 or V_2. Therefore, Equation (4.3) is still based on Equation (4.2), but instead of operating on the numbers of intercepting and non-intercepting sub-concepts, the equation makes use of the weights assigned in the semantic networks for the determination of the LSS.

Equation (4.3) can be demonstrated using a simple example based on Figure 4.6. The LSS for the values "Excavate" (V_1) and "Extract" (V_2) can be calculated using Equation (4.3) as follows:

$$\begin{aligned} \text{LSS}(V_1,V_2) &= (0.4 + 0.4 + 0.5 + 0.4 + 0.3) / [(0.1 + 0.3 + 0.1 + 0.4 + 0.3 + 0.1) + (0.4 \\ &\quad + 0.4 + 0.5 + 0.4 + 0.3)] \\ &= 2 / (2 + 3.3) = 0.606 \end{aligned}$$

The LSS for the values "Excavate" (V_1) and "Hack" (V_3) can also be calculated as follows:

$LSS(V_1,V_3)$ $= (0.5 + 0.4 + 0.3 + 0.1) / [(0.4 + 0.4 + 0.4 + 0.3 + 0.3 + 0.1) + (0.5 + 0.4 + 0.3 + 0.1)]$

$= 1.3 / (1.9 + 1.3) = 0.406$

The comparison above shows that "Excavate" is more similar to "Extract" than to "Hack", which is logical because both excavation and extraction usually involves moving of objects for the purpose of separating them from other objects. Even though, hacking also has the purpose of separating objects, it employs a (very) high frequency impact force to achieve that. So that, the hazards that hacking pose is generally different from that of excavation and extraction.

4.3.4 Global Similarity Score

The Global Similarity Score (GSS) is the overall similarity score between any two cases. During retrieval, the GSS is computed to determine the similarity between the input case and each of the stored cases. Most CBRSs compute GSS based on a weighted sum of the LSS of all the attribute-value pairs of the compared cases.

The following equation can be used for the computation of the GSS between two cases, C_1 and C_2,

$$GSS\ (C_1, C_2) = \sum(w_i \times LSS_i) / (\sum w_i) \qquad i = 1, 2, \ldots, n \qquad \textbf{(4.4)}$$

where w_i is the corresponding weight of attribute i, LSS_i the local similarity score for C_1's and C_2's values for the attribute i (refer to Equation (4.3)), and n the number of attributes.

For the SKMS, the attributes will refer to the six situational variables discussed in earlier sections.

The key issue in the calculation of the GSS is the determination of the weights for the different attributes, i.e. w_i. Various approaches have been adopted to establish the most suitable set of weights, but two of the most common approaches are, user-assignment and pre-assignment. The former approach is simply to allow users to assign the appropriate weights based on their assessment of the context of the problem. This approach is more appropriate if the users have the required level of expertise to make the necessary judgements. An example of the user-assignment approach can be found in Karim and Adeli (2003), where the Traffic Engineers using the CBRS are allowed to input the weights before the retrieval. On the other hand, pre-assignment approaches use fixed sets of weights that were determined based on different methods. For example, Chua et al. (2001) and Park and Han (2002) employed the Analytic Hierarchy Process (AHP) to determine the set of pre-assigned weights. Other methods to determine the pre-assigned set of weights also include optimisation methods like Genetic Algorithm (Zhang 1998) and Gradient Descent Method (Yau and Yang 1998). The pre-assignment approach can also be based on experts' subjective opinion. The advantage of the pre-assignment approach is that novice users need not assign the weights and focus only on providing the necessary information to describe the input case. However, this approach decreases the flexibility that might be needed to express the peculiarity of different input cases.

Due to the wide range of possible types of work situations and activities in construction projects, the importance of the situational variables would vary depending on the context of the case so that the pre-assignment approach appears to be impractical for the SKMS. It is very difficult to assign a fixed set of weights for such a large and varying domain. Thus, the user-assignment approach has been adopted for the SKMS. However, in contrast to the conventional user-assignment approach, where end users of the system assign the weights (or importance ratings), in the SKMS, the weights are assigned by incident investigators or risk assessment teams when they input the cases into the SKMS. This would allow the GSS calculated to better reflect the relevance of the incident sequence(s) to the input case. The proposed approach is feasible because incident investigators and risk assessment teams are usually required to record their findings and plans. Thus the assignment of importance rating will be operationally possible. Besides, the incident investigators and risk assessment teams should have sufficient expertise to assign the importance ratings.

The importance rating for each of the situational variables is stored in the "Impt" field of the table "tblSitVar" (see Figure 4.2). The rating is based on a Likert 5-point scale (see Table 4.1), where "1" would mean that the attribute-value pair is of low importance to the case, such as situational variables that provide only contextual information. On the other end of the scale, a "5" would indicate that the case's occurrence is highly dependent on that situational variable. The assignment of importance is also related to the selection of indices discussed in Section 4.2.2.2.

In contrast, for a conventional user-assigned approach it will be difficult for inexperienced users to assign the importance weights of the situational variables because they would not know the possible incident sequences that the situational variables might lead to. Even for experienced users the assignment of weights might be tedious and defeat the purpose of the SKMS, because they will have to first identify possible incident sequences and then assign the weights based on the number of incident events that is linked to each situational variable.

Table 4.1 Likert 5-point scale for assessment of importance of each necessary situational variable

Very Unimportant	Unimportant	Neither Important Nor Unimportant	Important	Very Important
1	2	3	4	5

4.3.5 Implementation of Case Retrieval in SKMS

Even though the SKMS retrieves both incident cases and past risk assessment trees to support risk assessment, both types of cases are retrieved using the retrieval process and the same LSS and GSS functions. The basic input for the retrieval process is the description of the current risk assessment context based on the set of situational variables described in Section 4.2. Each situational variable and its corresponding value are captured as an attribute-value pair. Hence, the input case is basically made up of a set of attribute-value pairs. As discussed earlier, only applicable situational variables need to be used and each situational variable can have more than one value. If the input case has more than one value for a particular situational variable, the value with the highest LSS will be used for the computation of the GSS. Chapter 6 will further illustrate how the GSS is computed.

In the prototype SKMS, incident cases with GSS greater than the user-specified GSS threshold, SS_{GT}, will be retrieved. For the prototype SKMS, the value of 0.6 is set as the default value for SS_{GT}. However, users can always increase the SS_{GT} to achieve higher relevance of retrieved cases or lower the SS_{GT} to ensure that there are more retrieved cases for utilisation. On the other hand, only the highest scoring risk assessment tree is retrieved and used in safety planning. The approach can be easily extended to retrieve more than one risk assessment tree and integrate all retrieved risk assessment trees using a similar methodology for the adaptation of incident cases.

Due to the novelty of the approach, available CBR shells are not able to facilitate the proposed approach. Thus, programming was done using Visual Basic for Applications (VBA) to create the necessary functions for LSS and GSS calculations. Details of the developed prototype will be further discussed in chapter 6.

4.4 Conclusions

This chapter discussed two of the most important and inter-dependent components of any CBRS, the knowledge representation and case retrieval components. The knowledge representation is structured based on two types of knowledge, the lessons learned and the context of the lessons learned or indices. The lessons learned were modelled as a "large case" with the snippets of the case being linked to the indices to allow for adaptation after retrieval. The indices correspond to the situational variables described in the Modified Loss Causation Model (MLCM) described in chapter 3. The proposed knowledge representation scheme is implemented in a relational database.

Case retrieval consists of two main parts, the determination of the Local Similarity Score (LSS) and the computation of the Global Similarity Score (GSS) based on the LSS. In the SKMS, the LSS is calculated using a weighted sub-attributes approach that is dependent on a series of semantic networks. The LSS are then combined through a weighting function to compute the Global Similarity Score (GSS). As opposed to the convention of having end users assigning the weights for the different attributes, the proposed approach adopted in the SKMS requires weights to be assigned during the input of the stored case, which allows for more appropriate assessment of importance of the situational variables.

Chapter 5

ADAPTATION AND UTILISATION OF RETRIEVED CASES

5.1 Introduction

Through the knowledge representation scheme and case retrieval mechanism, presented in the earlier chapter, a most similar risk assessment tree and a set of relevant incident cases has been retrieved to support the risk assessment process. Regardless of retrieval mechanism, retrieved cases always have the potential of containing irrelevant portions that should not be directly applied to the new case. In CBR, this problem is dealt with through various adaptation strategies and methods.

Kolodner (1993) proposed ten adaptation strategies, which are further classified under substitution, transformation and other methods. Substitution methods substitute the retrieved case's values with more appropriate values based on other relevant cases or some pre-determined model. On the other hand, transformation methods employ deletion or addition of relevant information to modify the retrieved lessons learned. Adaptation strategies classified under other methods are usually domain-specific and may be more complex, for example case-based adaptation, which retrieves and utilises the adaptation steps of past adaptation that were implemented (Vong et al. 2002).

It is possible to implement the various types of adaptation strategies proposed by Kolodner (1993) using some form of rules (Bergmann and Wilke 1998). For instance, the CBR Works development shell, Empolis Knowledge Management GmbH (2001) makes use of rules with a set of preconditions and corresponding conclusions to execute all adaptations. The set of preconditions are a conjunction of conditions that the retrieved case and/or input case must meet before the corresponding conclusions or list of actions is fired. Such rule-based adaptation is widely applied in different CBRS, for example Luu et al. (2003), Suh et al. (1998), and Sinha and May (2001).

However, it is noted that besides the types of adaptation strategies proposed by Kolodner (1993), the possible range of adaptation strategies is very wide. Adaptation is a very generic and high level process that can be implemented through any rational and sound method. For instance numerous CBRS have used different mathematical or statistical formulae for adaptation purposes (Chua et al. 2001; Suh et al. 1998). Other tools like genetic algorithm, decision tree, neural network, and statistical models can also be implemented to adapt the retrieved cases.

To meet the aims and suit the context of the SKMS, a novel adaptation strategy has been developed herein. The adaptation process is separated into two main steps: (1) adaptation during hazard identification, and (2) adaptation during risk analysis. Subsequent sections will present the adaptation process in detail.

5.2 Adaptation During Hazard Identification

Hazard identification is the process of identifying the "things" of a certain situation that can cause harm, who (or what) can be harmed and how the harm occurs (BSI 1996). In the context of the SKMS, the "situation" refers to the job step that the user is planning for and the "things" that can cause harm, or hazards, would refer specifically to each of the situational variables describing the job step, i.e. "action", "object-worked-on", "resource", "location", "nearby object", and "nearby action". The retrieved incident sequences provides information on how harm occurs, while information on who (or what) can be harmed are contained in consequences. However, for simplicity sake, the prototype SKMS has focused on severity of the harm and has omitted information on the possible victims or objects that can be damaged. This information can be easily included without any significant change in the design of the case base and the same principles and methods of implementation discussed herein still apply.

Even though the incident sequences of retrieved incident cases and risk assessment tree contain incident events relevant to the current context, there are portions of the retrieved cases which may not be applicable to the current situation and these must be removed. Furthermore, the retrieved incident cases will have to be incorporated into the risk assessment tree to produce a single adapted risk assessment tree that is relevant to the case in view.

The adaptation method used during SKMS's hazard identification belongs to the transformation type, but unlike classical transformation methods, the transformation adaptation employed is executed in two main steps. The first step focuses on the retrieved risk assessment tree, where the risk assessment tree is pruned to delete irrelevant incident events. The second step focuses on the set of retrieved incident cases, where the retrieved incident cases are trimmed to remove irrelevant incident events and then integrated into the risk assessment tree.

5.2.1 Adaptation of Retrieved Risk Assessment Tree

The adaptation of the retrieved risk assessment tree is based on the comparison of the indices of individual incident events in the risk assessment tree with the set of situational variables of the input case. The indices of incident events are subsets of the indices of the retrieved case. If any of the indices of an incident event is not similar (LSS < 0.6) to the corresponding situational variable of the input case, the incident event will be deleted.

An example of the proposed pruning adaptation is shown in Figure 5.1.

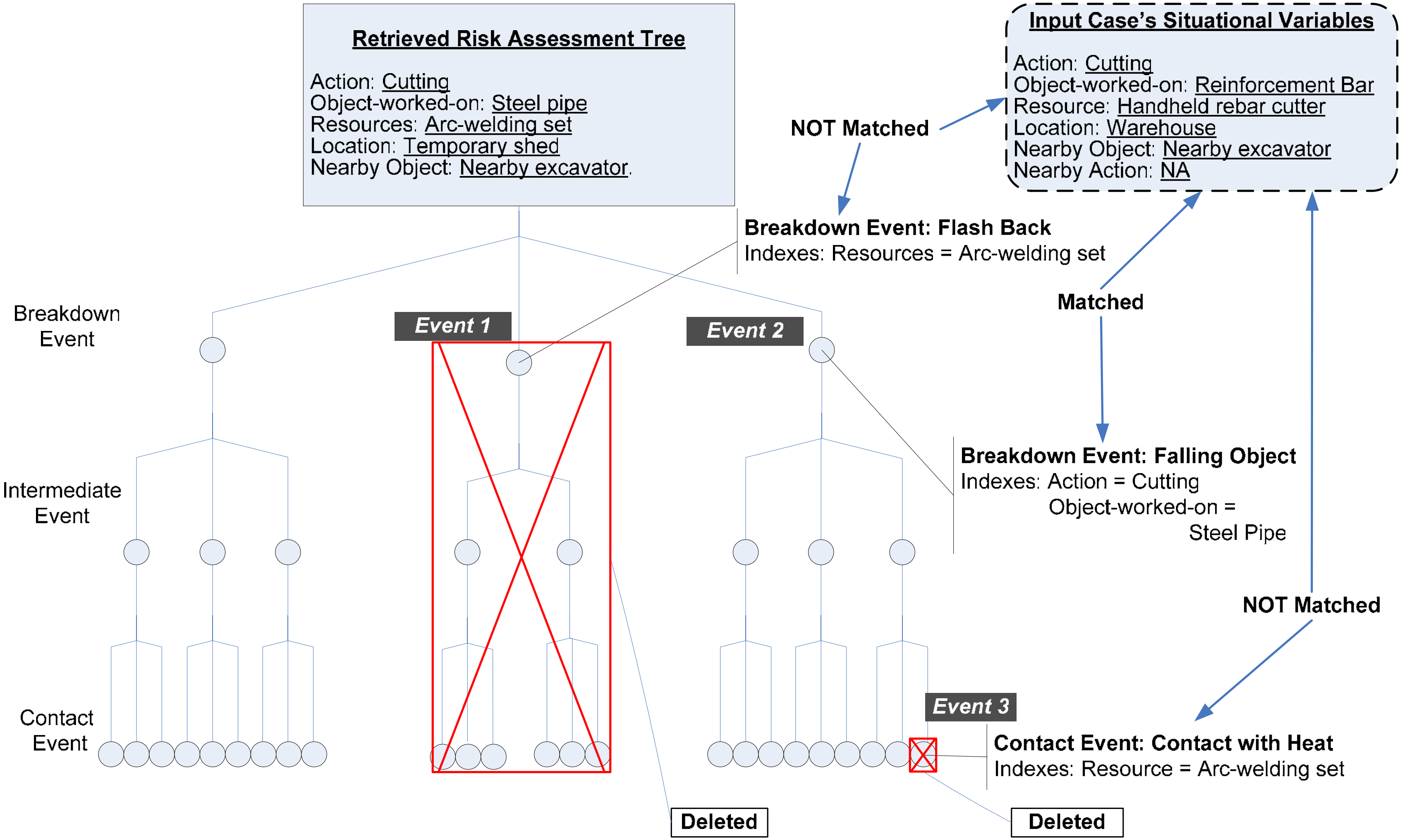


Figure 5.1 Example of a risk assessment tree being pruned

The inverted tree structure in Figure 5.1 represents the risk assessment tree that is being retrieved. The set of indices for the risk assessment tree is shown at the top of the inverted tree. The job step is a steel pipe cutting work that uses arc-welding equipment. The activity was executed in a temporary shed and near an excavator. For clarity sake, the consequences of each of the contact events are not presented in the risk assessment tree.

For illustration purposes, three events as highlighted in the figure are labelled for discussion. The indices of Events 1, 2 and 3 are presented in the different call-out boxes. Events 1 and 2 are breakdown events and Event 3 is a contact event. During adaptation, the relevance of incident events in the risk assessment tree is ascertained by determining the degree of match between indices of each corresponding event with the input case's situational variables.

As illustrated, Event 1 is pruned off because its Resource index, "Resource = Arc-welding set", is dissimilar (LSS = 0.1 < 0.6) to the corresponding index of the input case, i.e. "Resource = Handheld rebar cutter". As reflected in Figure 5.1, all events that follow Event 1 are also deleted. This is because each preceding event acts as a necessary condition for its following events. Thus once the preceding event is deleted, the subsequent events that are linked to it are also deleted.

On the other hand, the relevant Action and Object-worked-on indices of Event 2, "Action = Cutting" (LSS = 1) and "Object-worked-on = Steel Pipe" (LSS = 0.65), are similar (LSS ≥ 0.6) to the corresponding situational variables of the input case, "Action = Cutting" and "Object-worked-on = Reinforcement Bar" respectively. Thus, Event 2 is deemed to be a possible incident event in the new situation and is not deleted. Accordingly, Event 2 of the adapted tree will be updated with the corresponding situational variables of the input case, namely “Action = Cutting” and “Object-worked-on = Reinforcement Bar”.

Similar to Event 1, Event 3 was deleted because its Resource index, “Resource = Arc-welding set”, is dissimilar (LSS = 0.1 < 0.6) to the corresponding index of the input case, i.e. “Resource = Handheld rebar cutter”. The deletion was made despite its preceding event being accepted. This shows that even though the preceding event is a necessary condition for Event 3 to be accepted, it is not a sufficient condition that guarantees its acceptance.

As can be seen from the above example, only events with indices that matched (LSS $\geq$ 0.6) with the corresponding indices of the input case, for instance Event 2, will be accepted. In the example, Events 1 and 3 only have one index and the events were deleted when their only index was dissimilar to the corresponding index of the input case. In the situation when an event has more than one index, the event will also be deleted when any one of its indices failed to match with the corresponding index of the input case. Indices, like preceding events, are necessary conditions for an event to be accepted so that whenever one of these necessary conditions is not met the event will be deleted.

5.2.2 Adaptation of Retrieved Incident Cases

The adaptation of retrieved incident cases is executed after the retrieved risk assessment tree had been pruned. The adaptation is performed in two parts. The incident cases are first pruned to remove irrelevant incident events. Subsequently, incident events that are relevant but not identified earlier are inserted into the risk assessment tree.

The pruning of the incident cases is done in the same way as in the pruning of the risk assessment tree, but the insertion of the incident cases requires some additional checks. The incident event identified from the incident cases may have already been included in the retrieved risk assessment tree, and its insertion will cause duplication of incident sequences. Such duplication is not permissible because of the assumption of mutual exclusivity in the events for probabilistic analysis required in the risk analysis subsequently.

Although these duplicated events are not inserted, they hold important frequency or likelihood data that will be utilised during risk analysis using the adapted risk assessment tree (to be covered in Section 5.3). The link between the duplicated event and the corresponding event in the adapted risk assessment tree is maintained using the table "tblRelvEvent" as depicted in Figure 5.2. In this way, it is also possible to retrieve the past event or even the full incident case to provide justification for the adapted risk assessment tree.

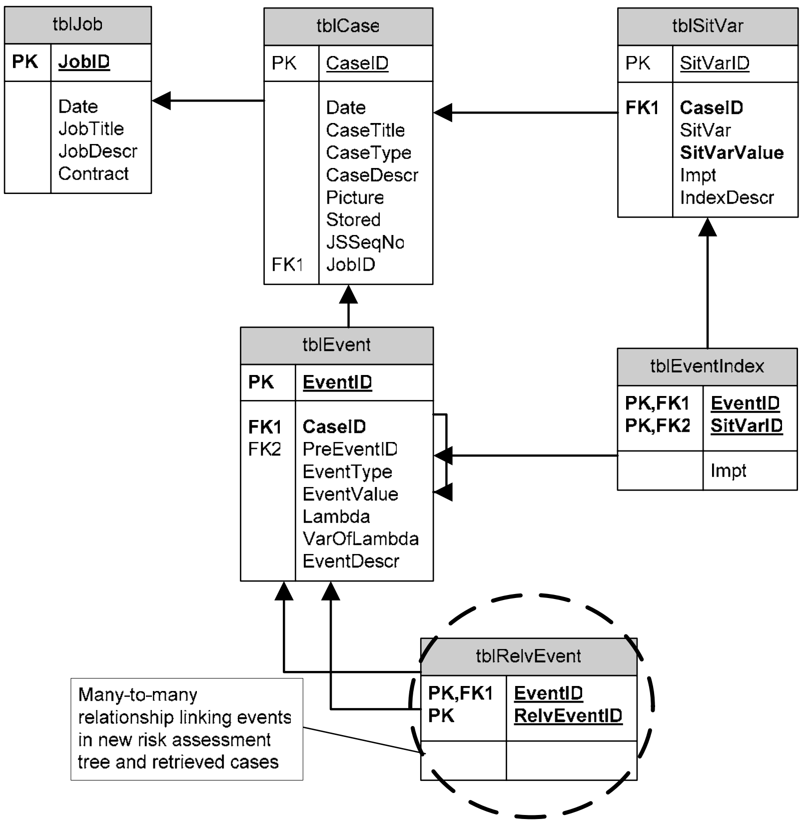


Figure 5.2 Table "tblRelvEvent" inserted to allow retrieval of likelihood data during risk analysis

The procedure for incorporating the incident cases into the risk assessment tree is illustrated with a hypothetical example depicted in Figure 5.3. The left hand side of Figure 5.3 shows a risk assessment tree that has been retrieved and pruned. On the right hand side are two incident cases, incidents 1 and 2, which have been retrieved based on similarity to the input case with situational variables shown at the top of the retrieved risk assessment tree.

Incidents 1 and 2 are based on actual incident cases obtained from the Land Transport Authority with a GSS of 1. All the incident events of the two incidents have the index "Resource = Pneumatic breaker", while the contact event of incident 1 has two other indices, "Action = Hack" and "Object-worked-on = Concrete column". Consequently, all the incident events are relevant to the input case (LSS = 1 ≥ 0.6 for all events), so that no pruning of incident events was necessary. In the case where the incident events are irrelevant, the pruning process will be based on the same procedure described in Section 5.2.1, by which the event nodes with LSS < 0.6 will be removed along with its child event nodes.

To insert the retrieved and relevant incident cases, all associated incident events are checked for duplication. An event is deemed to be duplicated when it occurs under the same preceding event (or root node for the case of breakdown events) and has the same event value. In the case of incident 2 (refer to Figure 5.3), when the system identifies that the breakdown event "Pneumatic breaker bounced off" is a duplicate, the breakdown event was not inserted. The system then continues to check the subsequent events of the duplicated event in the incident case. The subsequent events will still be inserted if they are not duplicated. In this example, the contact event of incident 2 was compared with the contact events under the duplicated breakdown event, "Pneumatic breaker bounced off", in the retrieved risk assessment

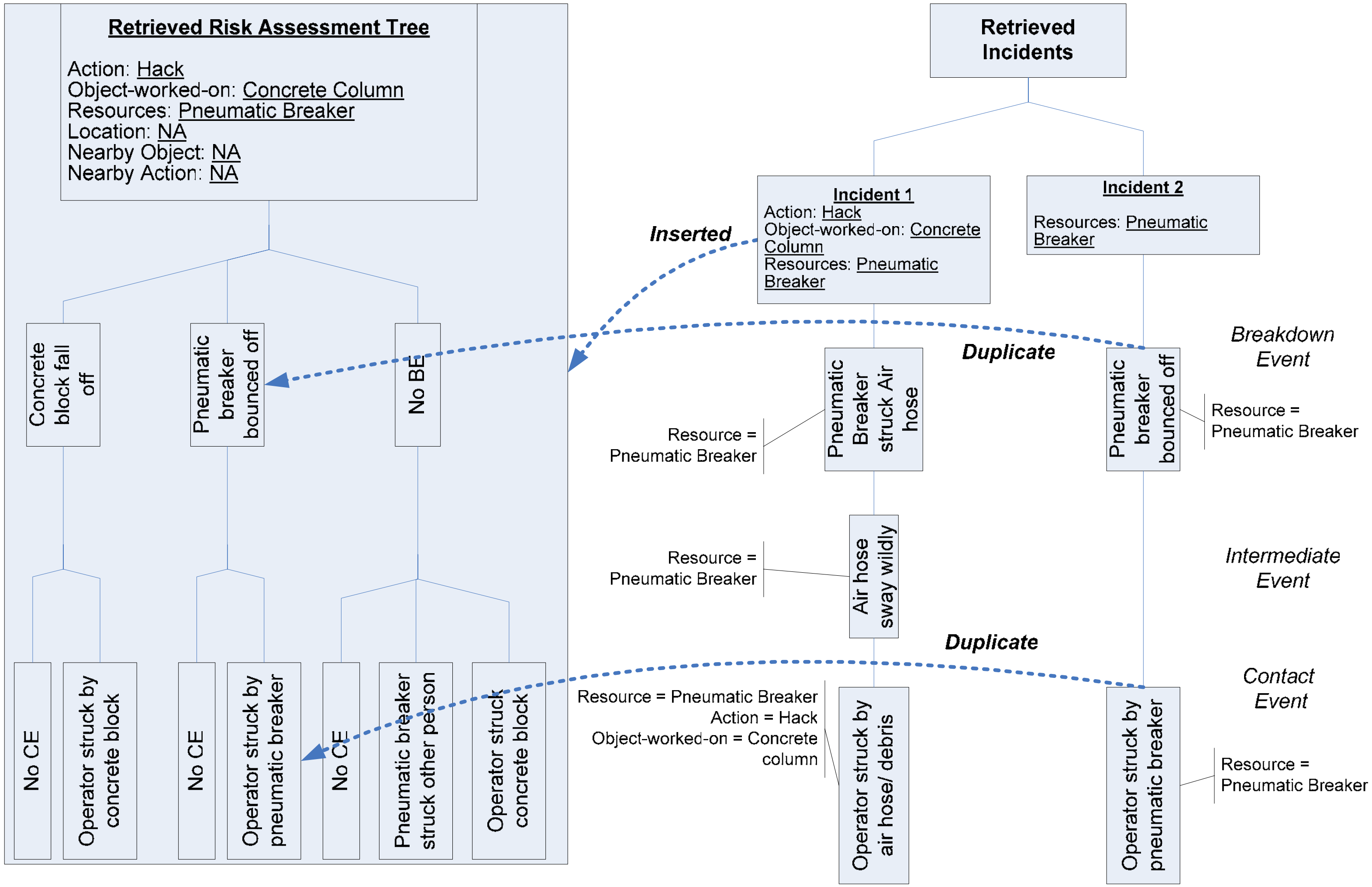


Figure 5.3 Example of incident cases being integrated in to a risk assessment tree

tree. Since the contact event, “Operator struck by pneumatic breaker”, is also duplicated, the contact event of incident 2 is not inserted.

5.3 Adaptation during Risk Analysis

Following hazard identification, risk analysis is carried out to assess the risk of the job step. *Risk* is generally defined as

$$Risk = P \bullet S \tag{5.1}$$

where P is the probability (or likelihood) of occurrence of “harm” and S is the expected severity of the “harm” (BSI, 1996, Baker et al., 1995, Redmill et al., 1999).

In the SKMS, risk analysis is conducted based on the risk assessment tree created during hazard identification. Figure 5.4 shows a simplified risk assessment tree with all the likelihood and severity values necessary to calculate the risk of the job step. At the top of the risk assessment tree is a set of situational variables that defines the job step. The parameter λ represents frequency of incident for the job step, and it has the unit number of incident per 50,000 man-hours worked (*no. per 50,000 mhr*). Furthermore, each of the incident events in the risk assessment tree has a conditional probability value, for example $P(B_1|SV)$, $P(C_{11}|SV,B_1)$ and $P(S_{113}|SV,B_1,C_{11})$, where SV refers to situational variables of the job step, B the breakdown event, C the contact event, and S the severity of the consequences of the preceding incident events. For clarity of presentation, intermediate events were not added into the Figure 5.4. During implementation, the intermediate events can be easily inserted with no significant changes in terms of principles and methodology.

The severity values are measured in terms of man-days lost (MDL), but any other quantitative measurements of severity, including subjective estimates can be used. As can be observed in Figure 5.4, severity can be classified into four categories. The four categories are based on Singapore's Land Transport Authority's (LTA) categorization, which are "None", "Minor", "Major" and "Fatal". The four categories corresponds to 0 MDL, 1-3 MDL, ≥ 4 MDL (but not fatal) and 6000 MDL respectively. It is noted that 6000 MDL is based on a common rule of thumb used in Singapore. LTA's categorization was based on Singapore's definition of reportable cases, where all incidents which results in more than 3 days medical leave or more than 24-hours hospitalization will have to be reported to the Ministry of Manpower (MOM).

Based on the risk assessment tree and the various likelihood parameters, the overall risk of the job step can be calculated as follows,

$$Risk(SV) = \lambda \bullet \{\sum_i P(B_i \mid SV) \bullet [\sum_j P(C_{ij} \mid SV, B_i) \bullet (\sum_k S_{ijk} \bullet P(S_{ijk} \mid SV, B_i, C_{ij}))]\} \tag{5.2}$$

where $(\sum_k P(S_{ijk} \mid SV, B_i, C_{ij}))$ the expected severity of the contact event C_{ij}, $[\sum_j P(C_{ij} \mid SV, B_i) \bullet (...)]$ the expected severity of the breakdown event B_i, and $\{\sum_i P(B_i \mid SV) \bullet [...]\}$ is the expected severity of the job step in the event that incidents occur. It can be observed that equation (5.2) is an expansion of equation (5.1).

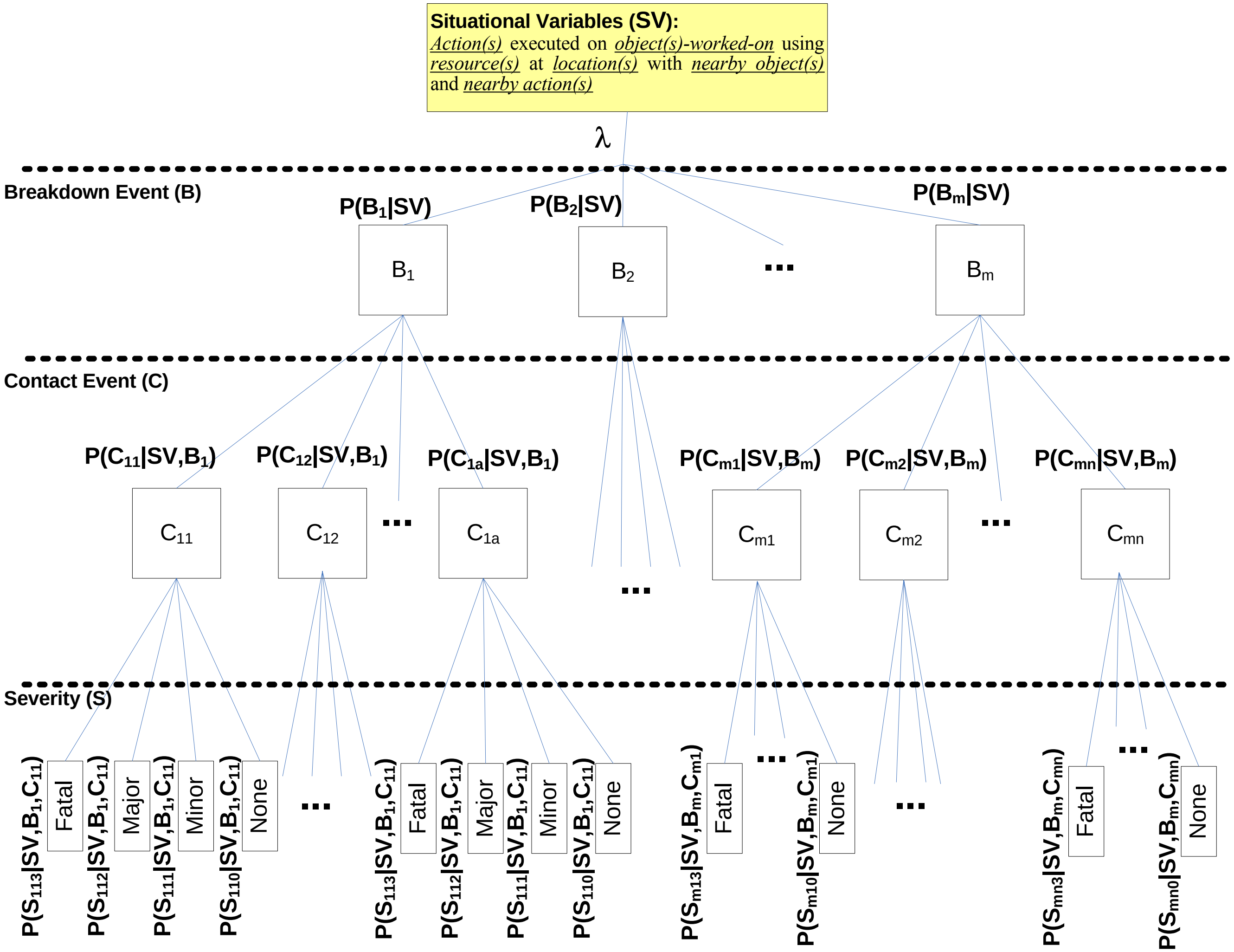


Figure 5.4 Risk assessment tree with various likelihood values

5.3.1 Adaptation for Estimation of Likelihood Values

In order to estimate the various likelihood values, two main sources of information, subjective and objective sources, are exploited. The SKMS retrieves both subjective and objective sources during retrieval. The most relevant risk assessment tree will usually contain subjective estimates provided by a previous risk assessment team. The subjective estimates can also be provided by the current risk assessment team and both current and previous subjective estimates can be averaged. These subjective sources will be based on the judgement, experience and knowledge of current or previous risk assessment team. However, subjective estimates had been known to be especially susceptible to various biases (Sanders and McCormick, 1992).

On the other hand, the set of relevant incident cases will contain objective likelihood estimates determined based on dates of incident occurrence and man-hours worked data of past projects. However, due to the fact that incidents are relatively rare events and that construction activities are seldom exactly the same, objective sources are often insufficient to produce statistically significant evaluations.

To overcome the above mentioned problems, the SKMS adapts the retrieved cases by integrating the subjective estimates from the retrieved risk assessment tree with the objective values from the set of relevant incident cases using the Bayesian approach (Ang and Tang 1975). The Bayesian approach is a systematic technique to incorporate objective data like incident occurrence data into subjective information, such as judgement, experience and intuition. The Bayesian approach is unconstrained by the size of incident data. Even in the event of no incidents, or having data based on a limited time frame, the approach can still be employed to "update" prior estimates. Furthermore, the Bayesian approach fits seamlessly into the proposed continual learning framework.

As opposed to classical statistics, the Bayesian method assumes that unknown parameters of a distribution are also random variables with their own distributions and the fundamental equation in the Bayesian approach is the simple but powerful Bayes' Theorem,

$$P(A|B)P(B) = P(B|A)P(A) \tag{5.3}$$

which will be elaborated in later sections. However, in order to apply the Bayesian approach, a suitable probability distribution function (PDF) based on a statistical model of construction incidents is adopted.

5.3.2 Statistical Model of Construction Incidents

The underlying assumption of a statistical model of construction incidents is that the occurrences of construction incidents are random. However, random does not refer to "without cause" or "unaffected by human actions", but instead it refers to the presence of variations. Variation means that two situations with similar characteristics will not guarantee the same outcome (Montgomery and Runger 1999). Randomness of incident occurrence has been highlighted in several incident causation models using different names like "window of accident opportunity", "chance" and "luck" (Ramsey 1985; Sanders and Shaw 1988; Reason 1990; McKinnon 2000). Hence, to allow the Modified Loss Causation Model (MLCM) to be employed as the basis for the statistical model an additional "chance" component had been inserted to reflect the random nature of incident occurrence (see Figure 5.5) (Chua and Goh 2004c).

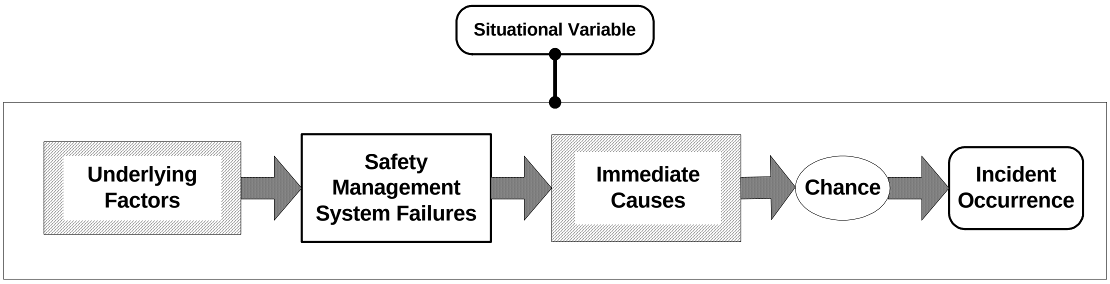


Figure 5.5 Simplified version of the Modified Loss Causation Model with an additional "Chance" component

Generally, modelling construction incident occurrences statistically will allow systematic characterization and analysis of the risk posed by construction incidents. A statistical approach presents a stable and sound foundation provided by mathematical boundaries and reasoning, thus improving the effectiveness of safety management. However, there has been a lack of formal studies to model construction incident occurrence statistically. Most of the past statistical studies on construction incidents (Jeong 1998; Cattledge et al., 1996; Kartam and Bouz 1998; Hinze, et al. 1998; Larsson and Field 2002) have been focused on summarising incident data obtained from different sources.

In contrast, traffic safety and reliability engineering researchers had utilised numerous probability distributions to model the occurrence of incidents in their respective areas, and one of the most commonly used is the Poisson distribution (Bendell 1991; Mordarres et al. 1999; Fridstrom, et al. 1995). However, because of differences in the environment, scale and nature of the processes, it would be prudent to verify that the distributions, in particular the Poisson distribution, are suitable for modelling the occurrence of construction incidents.

From a statistical point of view, the MLCM can be reorganised and interpreted as in Figure 5.6, wherein the model is now separated into two key components, namely the random component and the systematic component. The random component is inherent or objective, that is, the randomness is irremovable and uncontrollable. It is usually described by a probability density function (PDF),

$$f(\Phi, t) = P(X=x) \tag{5.4}$$

where Φ is a vector representing the parameters of the PDF, t the amount of exposure, for instance time or man-hours worked, X a random variable representing the number of incidents for t exposure, and x a specific value of X, e.g. a specific number of incidents for t exposure.

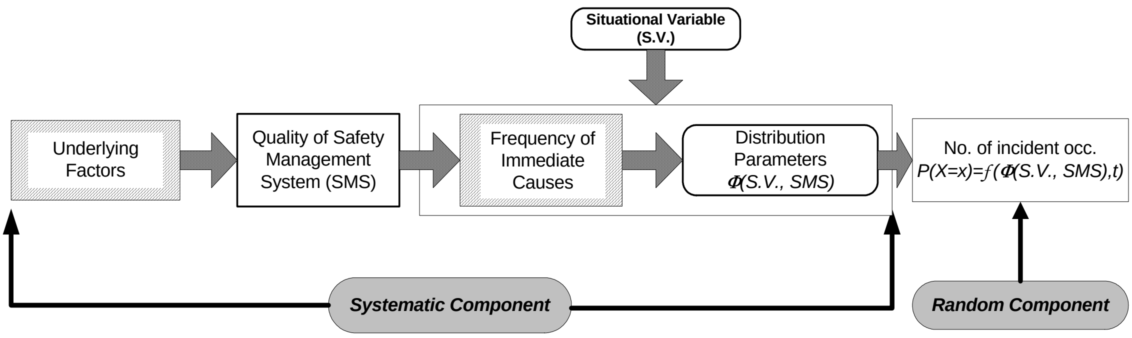


Figure 5.6 Statistical model of construction incident based on the MLCM

The systematic component, on the other hand, comprises the conditions or factors that are relatively controllable, and these would be the factors in the process system that influence the values of the parameters, Φ, of the PDF in Equation (5.4). As depicted in Figure 5.6, Φ is a vector of parameters that are influenced by a set of systematic independent variables. In the MLCM, these variables are broadly defined as situational variables and the quality of the SMS, which can be further categorised into more detailed variables. Even though immediate causes and underlying factors also influence Φ, they have been excluded and deemed to be represented by SMS quality, since they are highly correlated with the SMS quality. This will remove multi-collinearity problems in the model and retain the independence assumption required in many statistical methods. Moreover, among the three variables, quantitative measures of the SMS quality appear to be more readily available due to the rising trend of utilising quantitative SMS audit checklists in the construction industry.

The PDF (function $f(\cdot)$ in Equation (5.4)) that describes the random nature of construction incident occurrence is fundamental to the model. The choice of the PDF will define Φ, and hence the complexity of the statistical analyses. The range of possible PDFs is very wide, but ideally it should be simple to use and practical. On this count, the Poisson distribution is definitely one of the most preferred PDFs, as it only has one parameter. Its suitability to model the randomness of construction incident occurrence will be verified subsequently.

5.3.2.1 The Poisson Process Model

Based on the context of this research, the Poisson process can be considered as a type of counting process with the random variable $X(t)$ ($t \geq 0$), representing the total number of construction incidents that had occurred up to time t. In the broader sense, the variable t need not be time, but can be any continuum such as space and man-hours worked (number of workers × average number of hours worked). In comparison to conventional time intervals, man-hours worked will be able to better reflect the amount of activity on site, which in turn, directly reflects the risk exposure, and hence the probability of incident occurrence. Thus, man-hours worked has been used for the statistical model presented herein.

If construction incident occurrence follows a Poisson process, an interval of t man-hours worked can be partitioned into n number of subintervals of small enough length (t/n) such that there is at most one incident within each subinterval. This innocuous condition is necessary to facilitate the derivation of the Poisson distribution based on the binomial distribution, and it is a reasonable assumption in the construction context. For instance, an appropriate subinterval would be one man-minute worked or even one man-second worked, that is, the probability that there is more than one incident occurrence in an infinitesimally small interval is zero.

Another assumption in the Poisson process is the mutual independence of the number of incidents in disjointed intervals. This assumption is reasonable in the case of a construction project which is composed of many workers performing diverse activities at any time.

Consequently, in a Poisson distribution, the number of incident occurrences in an interval t, $X(t)$, with λ (> 0) as the mean number of incidents in t man-hours worked, can be represented by the probability mass function[1] (PMF)

$$f(x) = P(X(t) = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \tag{5.5}$$

[1] Generally, Probability Mass Function can be interpreted as the PDF of discrete random variables.

Equation (5.5) is based on the assumption that λ is constant. If this assumption is relaxed such that the probability of one incident in an interval is not constant, but a function of independent variables (Fridstrom et al., 1995) such as situational variables and SMS quality, then the PMF is modified as

$$f(x) = P(X(t) = x) = \frac{e^{-\lambda(SV,SMS)}[\lambda(SV,SMS)]^x}{x!} \qquad (5.6)$$

In this case, the distribution is known as the non-homogeneous Poisson distribution (Ross, 2000). With reference to the model depicted in Figure 5.6, function $f(\cdot)$ in the figure would refer to the non-homogeneous Poisson distribution in Equation (5.6), with $\lambda(SV,SMS)$ as the corresponding $\Phi(SV,SMS)$. It is noted that generally the SV and SMS of construction projects can vary as construction work progresses, but these variations are usually not significant. Hence, in most situation Equation (5.5), i.e. a homogeneous Poisson distribution, would be an adequate model.

The Poisson distribution was validated based on the incident data of fourteen contracts obtained from the LTA (Chua and Goh 2004c). Two goodness-of-fit tests, chi-square goodness-of-fit test (Conover 1980; Bendel, 1991) and the dispersion test (Cox and Lewis 1966; Nicholson 1985, 1986; Nicholson and Wong 1993), were applied on the data sets. The goodness-of-fit tests show that all the contracts can be modelled by the homogeneous Poisson distribution except for one of the contracts (Contract A), which requires the non-homogeneous Poisson distribution. It was identified that the failure of Contract A to fit the homogeneous Poisson distribution could be attributed to a change in the nature of work during the project or due to significant improvement in the SMS following the occurrence of earlier incidents.

Such major changes were not observed in the other 13 contracts. Thus, it can be generally accepted that the homogeneous Poisson distribution can be applied to model construction incident occurrence. The details of the analyses can be found in Appendix 4.

5.3.2.2 Partitioned Poisson Model

The earlier section shows that construction incident occurrence can be modelled by the homogeneous Poisson distribution. Thus, with the partitioned property of the Poisson process the distribution for categorized sub-processes can then be easily derived. Some possible categories include type of job step, type of incident or severity of incident. In general, it cannot be assumed that the distribution of incident occurrence for the categorized sub-processes will share the same distribution as the overall project, and even then, the parameters for the sub-processes are not readily available from the parameter for the overall project. The partitioned property of the Poisson model, however, implies that the sub-processes for various categorisations are also Poisson distributed and their parameters can be easily derived (Ross 2000; Wolff 1989). This is conditional on the assumption that the categorisation of these incidents is random. That is, the probability of an incident being classified as a category, say category “A”, is independent of its sequence, and categorisation of the preceding incident. Of particular interest to this research is the categorisation based on the type of job step, which is essentially defined by the set of situational variables depicted at the top of the inverted tree of Figure 5.4. Since the situational variables are low level descriptions of construction work, such that they are generally applicable to most possible types of site activities, it is reasonable to assume that the type of job step of an incident is random.

Figure 5.7 illustrates the concept of a partitioned Poisson process. The combined process can be visualised to be composed of several sub-processes defined by the possible values of a categorisation. The random variable X_{ij} is the number of

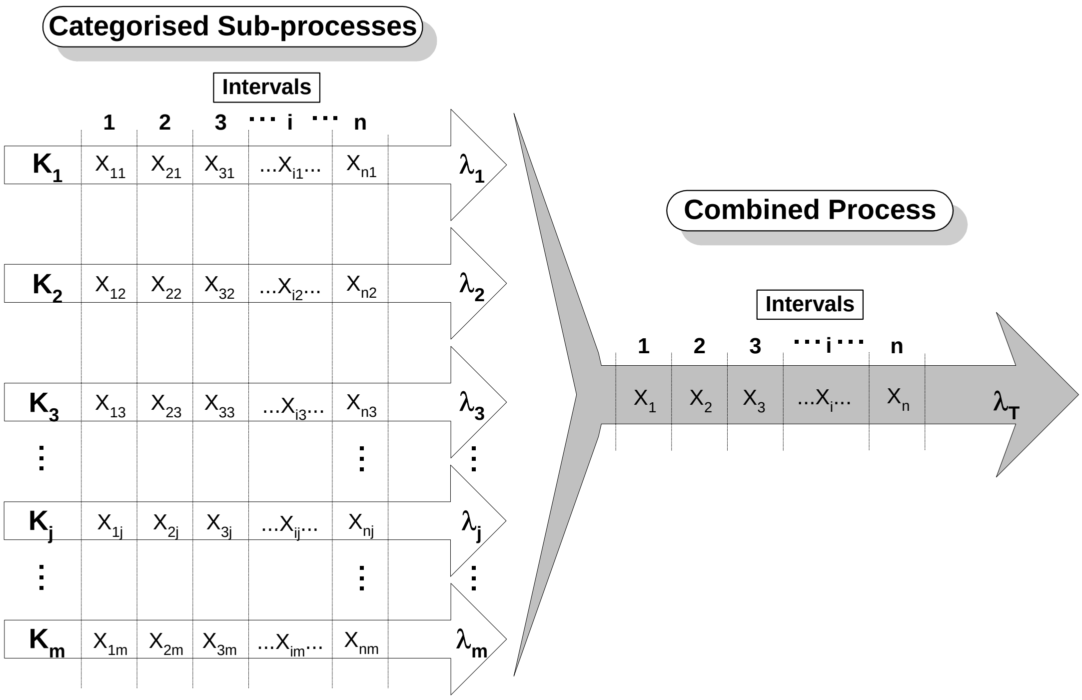


Figure 5.7 Partitioning a Poisson process into sub-processes

incident occurrences of sub-process j (j = 1 to m) in interval i (i = 1 to n). For an interval i, say i = 1, the total number of construction incident occurrence, X_1, is the sum of X_{1j} of all the sub-processes. If the estimated mean arrival rate of construction incidents for a project is $\overline{\lambda}_T$, the distribution of a particular sub-process j will also be Poisson distributed with mean arrival rate given by

$$\overline{\lambda}_j = P(E_j \mid E) \times \overline{\lambda}_T \tag{5.7}$$

where $P(E_j|E)$ is the probability of the categorisation given that an incident has occurred, E_j being the categorization of the sub-process j. $P(E_j|E)$ can be estimated as the relative frequency of E_j, i.e. number of incidents of category E_j/total number of incidents. For a mutually exclusive and collectively exhaustive categorization of the sub-processes,

$$\lambda_T = \sum_j \lambda_j \tag{5.8}$$

When applied to the context of this research, each project's incident occurrence process can be visualised to be made up of two sub-processes: (1) incidents that occurred in SV similar to the input case, and (2) incidents that occurred in SV not similar to the input case.

Furthermore, based on Equation (5.8) and the assumption that each of sub-processes is statistically independent, the variances of the rates of incident occurrence are related based on Equation (5.9),

$$Var(\lambda_T) = \sum_j Var(\lambda_j) \tag{5.9}$$

Equations (5.8) and (5.9) represent two simple, but important conditions that should be satisfied by the risk assessment tree. However, during hazard identification, incident events are pruned and inserted. Such alterations of the risk assessment tree cause inconsistency with Equations (5.8) and (5.9), because it implies removal and addition of sub-processes. The inconsistencies can be easily removed by adjusting the values of each process according to its sub-processes by applying Equations (5.8) and (5.9). The adjustment will be illustrated in the case study of the next chapter.

5.3.3 Bayesian Approach for Adaptation of Likelihood Values

Based on the establishment of the Poisson model of construction incidents and utilisation of the partitioned property of the Poisson process, it was derived that incident occurrence for each input case can be modelled by the Poisson distribution. Having determined the distribution of incident occurrence, it is then possible to implement the Bayesian approach to integrate subjective and objective likelihood or frequency values.

As mentioned earlier, the Bayesian approach treats the parameters of distribution as a random variable and hence each parameter will also have its own distribution. Accordingly, if $f'(\lambda)$ is the prior (initial estimated) distribution of the incident rate, λ, the posterior (revised) distribution $f''(\lambda)$ after incorporating incident observations may be obtained through Bayesian updating as

$$f''(\lambda) = kL(\lambda)f'(\lambda) \qquad \textbf{(5.10)}$$

where $L(\lambda)$ is the likelihood of observing the incident set assuming $f'(\lambda)$, and in which k is the normalising constant given by

$$k = [\int_{-\infty}^{\infty} L(\lambda)f'(\lambda)d\lambda]^{-1} \qquad \textbf{(5.11)}$$

Since construction incidents may be modelled as a Poisson process, it is convenient to assume a gamma distribution for λ in order to form a conjugate pair with the Poisson distribution (Ang and Tang 1975; Modarres et al. 1999). A conjugate pair will permit significant mathematical simplification to Equation (5.11) above, and both $f'(\lambda)$ and $f''(\lambda)$ will take the same gamma distribution but with different values for the shape parameter, κ, and scale parameter, $1/\upsilon$. In this way, if κ' and υ' are the corresponding prior estimates of the parameters, the revised parameters would be

$$\kappa'' = \kappa' + x \tag{5.12}$$

$$\upsilon'' = \upsilon' + t \tag{5.13}$$

where x is the number of incidents recorded in t intervals (of 50,000 *mhr*). Moreover, the parameters κ and υ of the gamma distribution are related to its mean and variance by the following relations:

$$\bar{\lambda} = \kappa/\upsilon \tag{5.14}$$

$$Var(\lambda) = \kappa/\upsilon^2 \tag{5.15}$$

Thus, the prior estimate of the mean rate of incident occurrence, $\bar{\lambda}'$, may be easily revised to $\bar{\lambda}''$ through the above relations using κ'' and υ''. However, as presented in Figure 5.6, λ is influenced by SMS quality and SV. Thus the information used for the estimation and updating of λ has to be of similar SMS quality and project level SV, e.g. project type. If the full statistical model is developed, more detailed adaptation can actually be developed to estimate and adapt $\overline{\lambda}$. At this stage, only information obtained from projects of similar SMS quality and project type are used in the Bayesian updating. It is noted that even though it is possible to include these variables during retrieval, inclusion of these variables might cause possible incident sequences to be eliminated, and hence not identified during hazard identification. Thus, it is more prudent to impose these conditions only during risk analysis.

The Bayesian approach can be illustrated with a hypothetical example based on a common construction scenario. Suppose a risk assessment team is conducting risk assessment for a precast column installation activity and one of the job steps has the following situational variables (SVs):

Lifting of *precast columns* using *tower crane and lifting gears* at *near structures* with *nearby concreting truck* and *concreting work*.

With this set of SVs for the input case, a risk assessment tree and a set of relevant incident cases can be retrieved. The retrieved cases are assumed to belong to projects with similar SMS quality and project type. In this way all the retrieved cases can be used to update the likelihood values.

Supposedly, during risk analysis, one of the incident sequences in the retrieved risk assessment tree, IS_1, is duplicated by the two retrieved incident cases, IC_a and IC_b , where IC_a and IC_b are retrieved from Project A, which is the only project in the case base. Figure 5.8 shows the occurrence of IC_a and IC_b on the project timeline (in 50,000 *mhr* intervals). As can be observed in Figure 5.8, the Poisson process 1 represents the incident occurrences of Project A. This Poisson process can be separated into two sub-processes, process 11 and 12, where process 12 represents incidents with similar SV as the input case of this example and having the same incident sequence as that of IS_1, while process 11 represents all other incident occurrences.

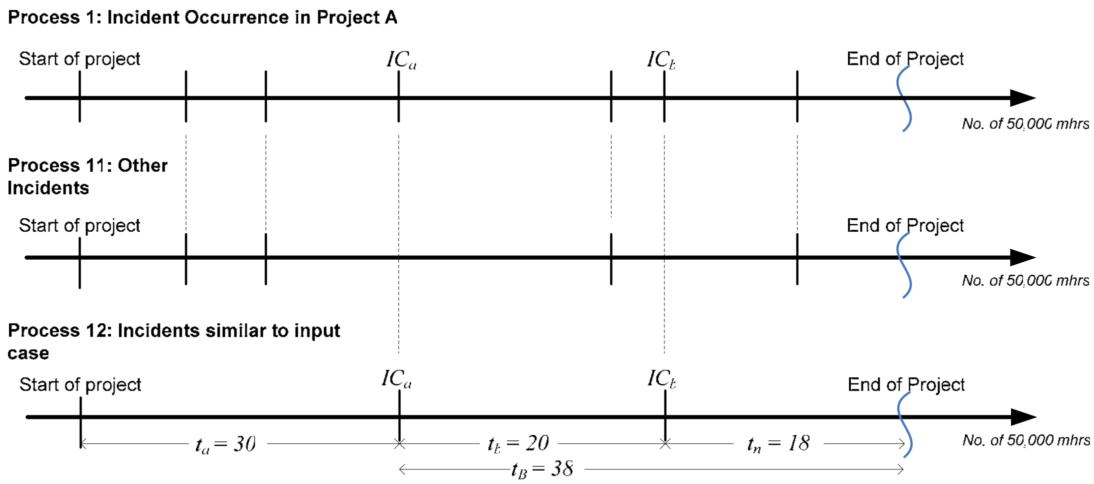


Figure 5.8 Poisson processes of Project A

Being the first occurrence in process 12, the t for IC_a, t_a, is simply the total *mhr* between the start of the project till the date of occurrence of IC_a. The t for subsequent incidents, such as IC_b, will be determined based on the *mhr* between occurrences. This is based on one of the assumptions of the Poisson process that any interval could be partitioned into small intervals that are independent (Montgomery and Runger 1999), such that the t for each event is independent of earlier t-values, and every time an event occurs the "Poisson clock" is set back to zero (Antelman 1997). However, due to the fact that IC_b, is the last occurrence in process 12, t_B ($t_B = t_b + t_n$) instead of t_b (see Figure 5.8) is recorded as its t-value. This is to account for the non-occurrence of incidents between the occurrence of IC_b and the end of the project. This approach is also consistent with concept of Poisson sampling (Antelman 1997), where the period of observation, T, is fixed. The fixed T implies that the period of non-occurrence just before the end of observation is also taken into consideration.

The t-values of the incident cases are then used to update the frequency values attached to the relevant leaf nodes of the risk assessment tree. The update is applied on the leaf nodes because in a risk assessment tree the leaf nodes' frequency values describe the frequency of occurrence of the intersection of the leaf node and all its higher nodes, i.e. the whole incident sequence. In this example, suppose that the prior likelihood estimate is $\bar{\lambda}' = 0.01$ per 50,000 *mhr*, and the coefficient of variance (C.O.V.) is 33%.

For this example, say $t_a = 30$, $t_b = 20$, $t_n = 18$ and $t_B = t_b + t_n = 38$ (see Figure 5.8). The number of incident cases would correspond to x (=2) in Equation (5.12), while $t_a + t_B$ (= 68) or the total project *mhr* of the case base would correspond to t of Equation (5.13). With the above information, it is then possible to derive the updated or posterior mean arrival rate, $\bar{\lambda}''$, as follows:

Using Equations (5.14) and (5.15), C.O.V. = $\sqrt{Var(\lambda)}/\bar{\lambda} = 0.33$

$$\sqrt{\kappa'/\upsilon'^2}/(\kappa'/\upsilon') = 0.33$$

$$1/\sqrt{\kappa'} = 0.33$$

$$\kappa' = 9.183$$

Substituting $\kappa' = 9.183$ and $\bar{\lambda}' = 0.01$ into Equation (5.14) would yield,

$$\upsilon' = \kappa'/\bar{\lambda}' = 918.3$$

Substituting $\kappa' = 9.183$, $\upsilon' = 918.3$, $x = 2$, and $t = 68$ into Equations (5.12) and (5.13), the revised parameters would be

$$\kappa'' = \kappa' + x = 9.183 + 2 = 11.183$$

$$\upsilon'' = \upsilon' + t = 918.3 + 68 = 986.3$$

Therefore from Equation (5.14) and (5.15), $\bar{\lambda}'' = \kappa''/\upsilon'' = 0.01134$, updated $Var(\lambda) = \kappa''/\upsilon''^2 = 1.150 \times 10^{-05}$ and C.O.V. = 29.9%. As can be observed $\bar{\lambda}''$ is slightly higher than $\bar{\lambda}'$ and the C.O.V. had decreased. This is because the objective data has a frequency rate higher than the subjective estimates, and the Bayesian approach allows the two estimates to be integrated systematically so as to achieve a balance. Furthermore, the reduction in C.O.V. shows that the uncertainty in the prior estimates is decreased due to the inclusion of objective data.

In the event that the case base has multiple projects with the same SMS quality and project type as the input case, the x and t values for a particular branch of the risk assessment tree will be based on all projects. x (Equation (5.12)) will be the total number of retrieved incident cases (GSS $\geq$ 0.6) with the same incident sequence and t (Equation (5.13)) will be the total *mhr* of all the relevant projects. This means that the non-occurrence (x = 0) of a particular incident sequence in a project will also be accounted for in the adaptation.

All the leaf nodes of the risk assessment tree can also be updated in the same manner as shown above. Once all the leaf nodes of the risk assessment tree have been updated, the frequency values and variances of the rest of the tree can then be easily derived based on Equations (5.8) and (5.9). Each node's frequency value and variance are updated based on the summation of the frequency values and variances of its child nodes respectively. In this way the updating will be done in a bottom-up manner. The bottom-up propagation will be further illustrated in chapter 6.

Once all the frequency values of a risk assessment tree are adapted, Equation (5.2) can then be used to ascertain the level of risk of the job step. The risks posed by different incident events can also be determined by aggregating the risk values of incident events contained in the corresponding sub-trees under the relevant incident event. These lower level risks can then assist the risk assessment team in identifying the higher risk incident events for more focused risk control selection. A case study will be illustrated in the next chapter to validate and further clarify the proposed approach.

5.4 Conclusions

This chapter has presented important adaptation concepts to facilitate the risk assessment process. The adaptation strategy is separated into two main parts to facilitate hazard identification and risk analysis sequentially. During hazard identification the focus is on facilitating hazard identification by ensuring that irrelevant incident events of the retrieved risk assessment tree and incident cases are pruned off. Furthermore, an approach has been devised to integrate the incident cases and risk assessment tree. During the integration, special attention is paid to the careful retention of frequency data of retrieved cases so that they can be utilised for the adaptation of likelihood values for risk analysis purposes.

After hazard identification, an adapted risk assessment tree is produced, and the next task is to estimate all the likelihood values attached to each of the incident event. The estimation is based on the Bayesian approach, which allows both subjective data found in the retrieved risk assessment and the objective data found in the retrieved incident cases to be integrated and produce a more balanced estimation. However, in order to apply the Bayesian approach a statistical interpretation of construction incidents is warranted. Thus, a Poisson model of incident occurrence was validated using goodness-of-fit tests applied on actual incident data. The simple, but powerful, Poisson model provides the basis for the proposed adaptation approach of frequency values, which basically uses subjective estimates of the frequency values as the prior estimates and these estimates are integrated systematically with relevant objective data to produced more balanced posterior estimates.

As in the case of retrieval functions (chapter 4), the programming language Visual Basic for Applications (VBA), which is part of Microsoft Access, was used to create the necessary functions to implement the adaptation processes presented.

Chapter 6

VALIDATION CASE STUDY

6.1 Introduction

This chapter presents a case study to validate the key concepts presented in the earlier chapters. The validation is based on the demonstration of how a risk assessment tree can be constructed based on past experiences and hence achieving the desired feedback and learning process proposed. In the case study, the prototype SKMS will show how cases modelled in the proposed knowledge representation framework can be retrieved and adapted to produce a final risk assessment plan that facilitates identification of possible hazards and analysis of the corresponding risks.

6.2 Case Base for Case Study

The case base utilised in the case study consists of two types of cases: incident cases and risk assessment trees. The incident cases were obtained from the Land Transport Authority's (LTA) Safety Information System (SITS). All the cases belong to the same project and to ensure the integrity of the cases, the details of the incident cases were verified through interviews with relevant site personnel and review of appropriate site documents. It is noted that the number of cases does not affect the essence of the case study because the key purpose is to validate the feasibility of the proposed concepts and methodologies. Furthermore, all CBRS are learning systems that accumulate knowledge as more experiences are being accumulated and are capable of fully utilising available knowledge. Thus the case study also demonstrates how a "young" CBRS can still produce valid and useful risk assessment plans.

The incident cases in the case base belong to a railway project and it involves construction of precast viaducts and above ground train stations. The contract has a total of 59 reported incidents after 1,444,8300 *mhr* of work (project duration of 3 years and 9 months). Figure 6.1 shows the distribution of the incident cases in terms of severity measured by number of man-days lost. As can be observed, most of the incident cases (39 cases or 66.1%) are less than 3 MDL and are not required to be reported to the authorities. There are 20 cases (33.9%) which are required by law to be reported to the Ministry of Manpower (MOM). Among these MOM reportable cases there is a case with fatality, where one worker was killed. The incident cases occurred in a wide variety of types of activity, such as soil boring, hoisting, concreting and manual handling work. Such variety is advantageous because it forms a rich source of knowledge.

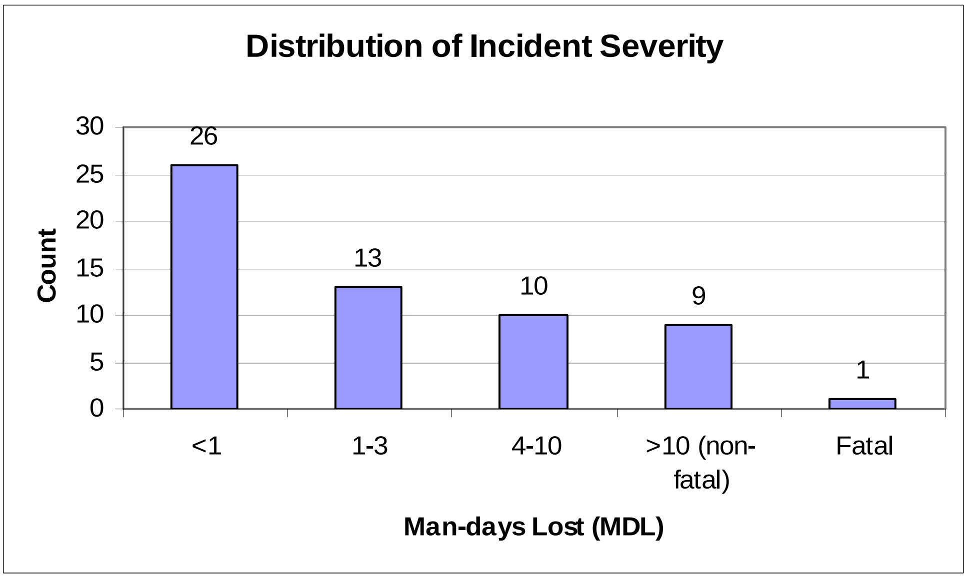


Figure 6.1 Distribution of incident severity in terms of man-days lost

Besides the 59 incident cases, the case base also contains ten risk assessment trees (see Table 6.1). The risk assessment trees are based on the safety documents obtained from numerous sources, which included main contractors' safety management systems, tender documents submitted to LTA, and training materials of Mine Safety and Health Administration (MSHA 2004). The knowledge representation process was carried out based on the representation scheme proposed in chapter 4 and in consultation with two experienced construction safety practitioners with at least 8 years of construction safety experience each. The safety experts were asked to verify the content of the cases and provide inputs like assignment of subjective likelihood values and appropriate indexes whenever necessary.

Table 6.1 Case titles of risk assessment trees in case base

No.	Case Title
1	Gas-cutting in confined space (tank)
2	Rigging up precast element
3	Lift precast wall using crawler crane
4	Arc welding of suspended pipes in trench
5	Concreting work using bucket
6	Lowering pipe into trench using excavator
7	Loading truck with soil using excavator
8	Concrete breaking
9	Frame scaffold erection
10	Gas-cutting of H-pile

6.3 Case Study

The job scenario that is being assessed in the case study is based on a common site activity as shown in Figure 6.2. The picture (Figure 6.2) was taken on an actual construction site in Singapore and it shows part of a material delivery activity where the lorry crane is in the process of unloading a bundle of timber strips. The case study will show how a risk assessment tree is created for this work activity based on the available knowledge stored in the incident cases and risk assessment trees.

Figure 6.2 Scenario for the risk assessment case study

6.3.1 Case Retrieval

Figure 6.3 shows the graphical user interface (GUI) in the prototype SKMS that is used for the input and viewing of both cases stored in the case base (stored cases) and the current case (input case). As shown in Figure 6.3, the top portion of the GUI contains details of the current case and there are also several command buttons that executes the retrieval and adaptation functions. The bottom portion of the GUI contains a subform that would contain the incident events of the case. These incident events can be easily transferred to graphical and visualisation tools, such as Microsoft Visio, to create a more visual form of the incident sequences.

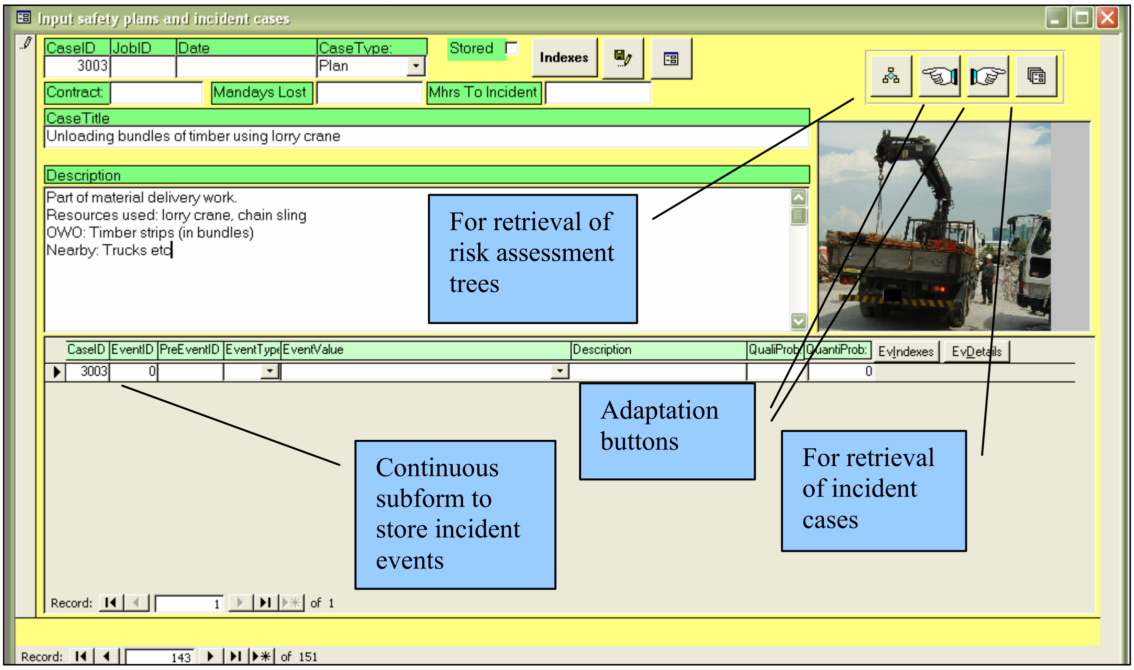
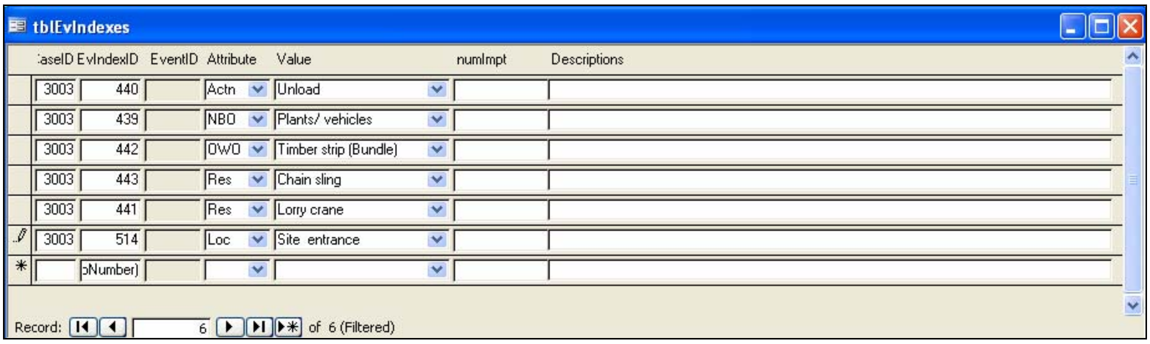


Figure 6.3 Graphical user interface for both input and stored cases

Figure 6.4 shows the GUI used to store the situational variables and indexes of stored and input cases. As discussed in chapter 4, incident investigators and risk assessment teams will assign weights to reflect relative importance of different situational variables of a stored case in the field "numImpt". These weights are used for the calculation of global similarity scores between the input case and each stored case in the case base. End users using the system to carry out risk assessment will not be required to assign the weights for the input case.



tblEvIndexes

:aseID	EvIndexID	EventID	Attribute	Value	numImpt	Descriptions
3003	440		Actn	Unload		
3003	439		NBO	Plants/ vehicles		
3003	442		OWO	Timber strip (Bundle)		
3003	443		Res	Chain sling		
3003	441		Res	Lorry crane		
3003	514		Loc	Site entrance		
	oNumber)					

Record: 6 of 6 (Filtered)

Figure 6.4 Graphical user interface for situational variables and indexes

For this case study, the input situational variables are shown in Figure 6.4. The set of situational variables is based on the following description of the job step, "*Unloading* of *Timber strip (Bundle)* using *Chain sling* and *Lorry crane* at *Site entrance* with nearby *Plants/ vehicles*."

Table 6.2 shows the global similarity scores (GSS) between the input case and each of the 10 risk assessment trees stored in the case base. Intuitively from the descriptions of the case, the risk assessment trees with GSS>0.5 is more similar to the input case as compared to risk assessment trees with GSS<0.5. Table 6.3 shows a detailed breakdown of the local similarity scores (LSS) between the input case and the most similar (i.e. retrieved) risk assessment tree. The input case and the retrieved case are very similar in terms of action, nearby object and one of the resources used (Excavator and Lorry Crane). The high GSS is contributed by the high weights for the situational variables with high LSS. It is highlighted that case number 3 also has a relatively high GSS, and it is possible that both cases 6 and 3 are relevant to the input case. This point will be further discussed in Section 6.4.

Table 6.2 The global similarity scores of all risk assessment trees in the case base

No.	Case Title	GSS
6	Lowering pipe into trench using excavator	0.61
3	Lift precast wall using crawler crane	0.59
7	Loading truck with soil using excavator	0.51
5	Concreting work using bucket	0.39
2	Rigging up precast element	0.27
4	Arc welding of suspended pipes in trench	0.24
8	Concrete breaking	0.15
9	Frame scaffold erection	0.10
10	Gas-cutting of H-pile	0.10
1	Gas-cutting in confined space (tank)	0.05

Table 6.3 Local similarity scores of retrieved risk assessment tree

Situational Variables	Value		LSS	Weights	Weighted LSS
	Stored Case	Input Case			
Action	Lower	Unload	0.73	5	3.64
Location	Trench	Site entrance	0.00	2	0.00
Nearby Object	Plants/ vehicles	Plants/ vehicles	1.00	3	3.00
Object-worked-on	Pipe	Timber strip (Bundle)	0.11	2	0.21
Resource	Lifting gear	Chain sling	0.60	2	1.20
Resource	Excavator	Lorry crane	0.80	3	2.40
			Total:	17	10.45

As mentioned in chapter 4, when the input case has more than one value for a particular situational variable, the highest LSS value will be used for the calculation of the GSS. In this case study, the input case has two resources, "Chain sling" and "Lorry crane". Thus, when the index of the stored case, "Resource = Excavator", is compared with the two Resource indexes of the input case, "Resource = Chain Sling" and "Resource = Lorry crane", the corresponding LSS are 0.09 and 0.80 respectively. Therefore, the index with higher LSS, i.e. "Lorry crane", is used to match against the stored case's "Excavator". Similarly, for the stored case's other Resource index, "Resource = Lifting gear", the chain sling with an LSS of 0.6 is used for the matching.

Figure 6.5 shows the distribution of the GSS of all the incident cases in the case base. As can be observed, the majority of the stored cases, or 32 (54%) cases, have GSS between 0 and 0.2. Besides that, another 22 cases (37%) have GSS between 0.2 and 0.6. In contrast, only 5 cases or 8.5% has GSS $\geq$ 0.6. Table 6.4 shows the GSS and LSS of the incident cases with GSS $\geq$ 0.6. These cases were returned for adaptation. Cases 1153 and 141 occurred during lifting work using mobile plants, case 230 occurred during unloading of material, and cases 160 and 292 are related to mobile plants that were used during construction activities.

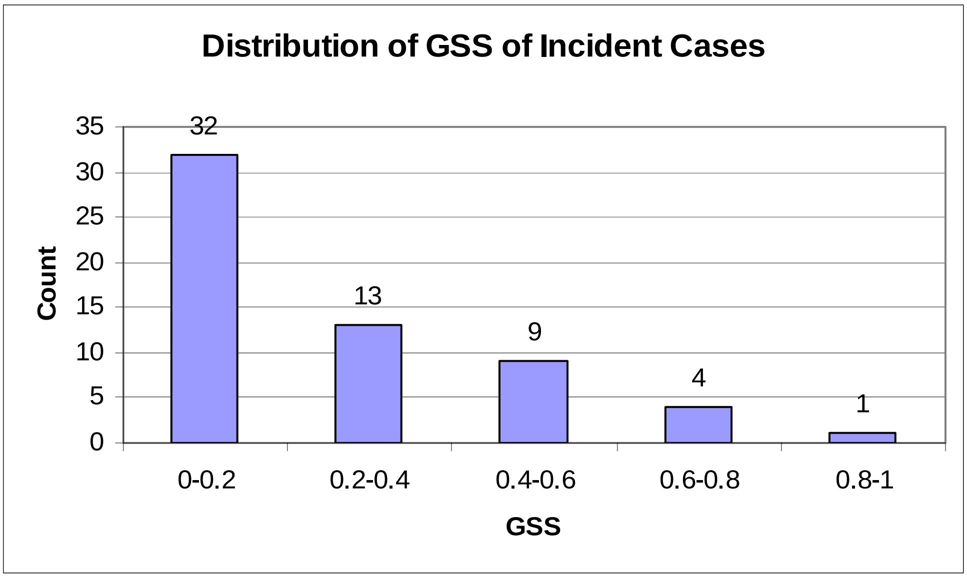


Figure 6.5 Distribution of GSS of incident cases

Table 6.4 Local similarity scores of retrieved incident cases

No.	GSS	Case Title	Situational Variables	Values: Stored Case	Values: Input Case	LSS	Weights	Weighted LSS
160	0.86	Worker injured by forklift	Resource	Forklift	Lorry crane	0.86	5	4.32
230	0.78	Finger trapped between I-beams during unloading	Action	Unload	Unload	1.00	5	5.00
			Object-worked-on	I-beams	Timber strip (Bundle)	0.18	3	0.53
			Resource	Lorry crane	Lorry crane	1.00	3	3.00
292	0.77	Worker injured while coming down from lorry	Resource	Lorry	Lorry crane	0.77	5	3.86
1153	0.74	Parapet wall fall from height while lifting	Action	Lift	Unload	0.73	4	2.91
			Object-worked-on	Precast parapet wall	Timber strip (Bundle)	0.07	1	0.07
			Resource	Crane	Lorry crane	0.64	3	1.91
			Resource	Chain sling	Chain sling	1.00	4	4.00
141	0.67	Worker injured by I-beam during lifting	Action	Lift	Unload	0.73	4	2.91
			Object-worked-on	I-beam	Timber strip (Bundle)	0.20	2	0.40
			Resource	Mobile crane	Lorry crane	0.91	3	2.73

6.3.2 Hazard Identification

During hazard identification, the key aim is to identify possible incident sequences that can occur during the execution of the job step described in the input case. The retrieved risk assessment tree and incident cases will first be adapted by pruning off irrelevant incident events. Following that the incident cases and risk assessment tree will be integrated.

Figure 6.6 shows the retrieved risk assessment tree being pruned (the critical index with the lowest LSS for each event is noted). A total of seven incident events (breakdown, intermediate and contact events) were being removed. The breakdown events (BEs), “Soil structure collapse” (Event 221) and “Soil/ objects fall into trench/ excavation” (Event 224) were deleted because one of the indexes of these BEs, “Location = Trench” (index (3)), is not similar (minimum LSS = 0) to the Location index of the input case “Location = Site entrance” (index (c)). Since the events following the two BEs were conditional on the occurrence of the BEs, these subsequent events were also deleted. The contact event (CE), “Cut by lifted object” (Event 195) is also deleted because one of the event’s indexes (index (2)), “Object-worked-on (OWO) = Pipe” is dissimilar (minimum LSS = 0.11 < 0.6) to the OWO of the input case (index (b)), “Timber strip (Bundle)”. The rest of the incident events were not removed because the minimum LSS is at least 0.6 or higher.

The set of five retrieved incident cases of Table 6.4 are shown in Figure 6.7. None of the incident cases need to be pruned because all the incident events’ indexes are matched by the input case’s indexes, i.e. minimum LSS ≥ 0.6. The incident cases were then checked for duplication with the incident sequences in the adapted risk assessment tree. As indicated in Figure 6.7, five of the incident events were duplicated (not including consequence events), while the remaining five were not identified in

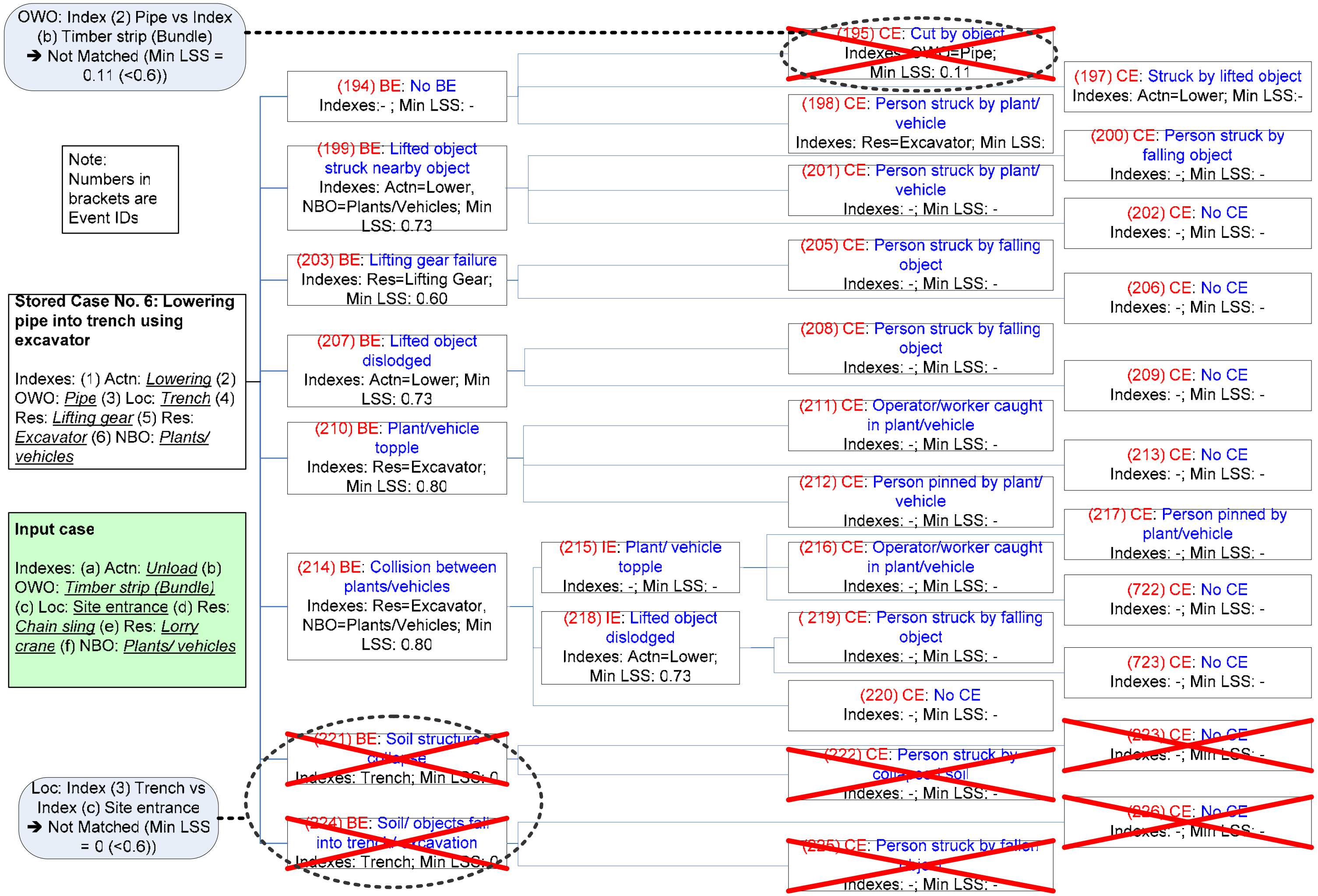


Figure 6.6 Retrieved risk assessment tree being pruned

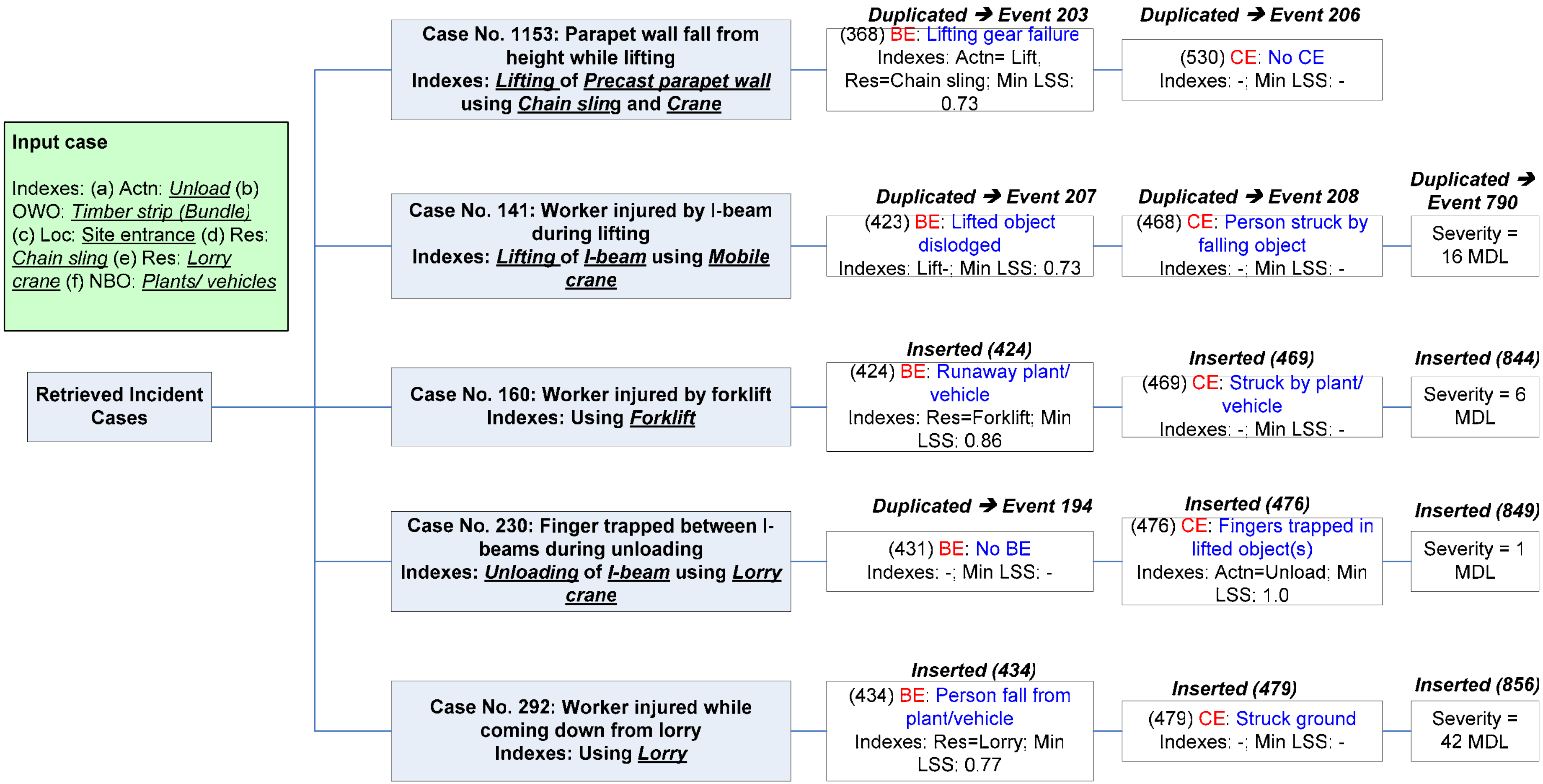


Figure 6.7 Incident sequences of the retrieved incident cases

the risk assessment tree and were hence inserted. Links to the duplicated events were maintained to ensure that their frequency can still be retrieved during risk analysis.

The risk assessment tree containing the full set of identified incident sequences is shown in Figure 6.8. The shaded incident events are based on the retrieved incident cases. All of the contact events also has a set of consequence events described by five severity categories in terms of man-days lost (MDL): (1) < 1 MDL, (2) 1-3 MDL, (3) 4-10 MDL, (4) > 10 MDL, and (5) fatal (F). The consequence events are essential for the risk analysis described in the next section.

6.3.3 Risk Analysis

The SKMS risk analysis process aims to estimate the likelihood or frequency values for each of the incident event in the adapted risk assessment tree (see Figure 6.8) based on both subjective and objective sources. The Bayesian approach is used to integrate the subjective and objective estimates to arrive at a more balanced estimation. However, prior to the Bayesian updating, the likelihood values in the retrieved risk assessment tree have to be fine-tuned to account for the adaptation during hazard identification.

As noted in chapter 5, the retrieved cases that are used for risk analysis should have similar SMS quality and project type as that of the input case. This is to ensure the relevance of the likelihood values in the cases. In the case study, the SMS quality and project types of the input case and stored cases are deliberately similar to simplify the case study. Since the incident cases were obtained from a railway construction project with an above average SMS, the input case is assumed to have the same characteristics. Similarly, during the assignment of likelihood values for the stored risk assessment trees, the safety experts were asked to assume the same context. In this way, all the likelihood values can be utilised during risk analysis. The frequency

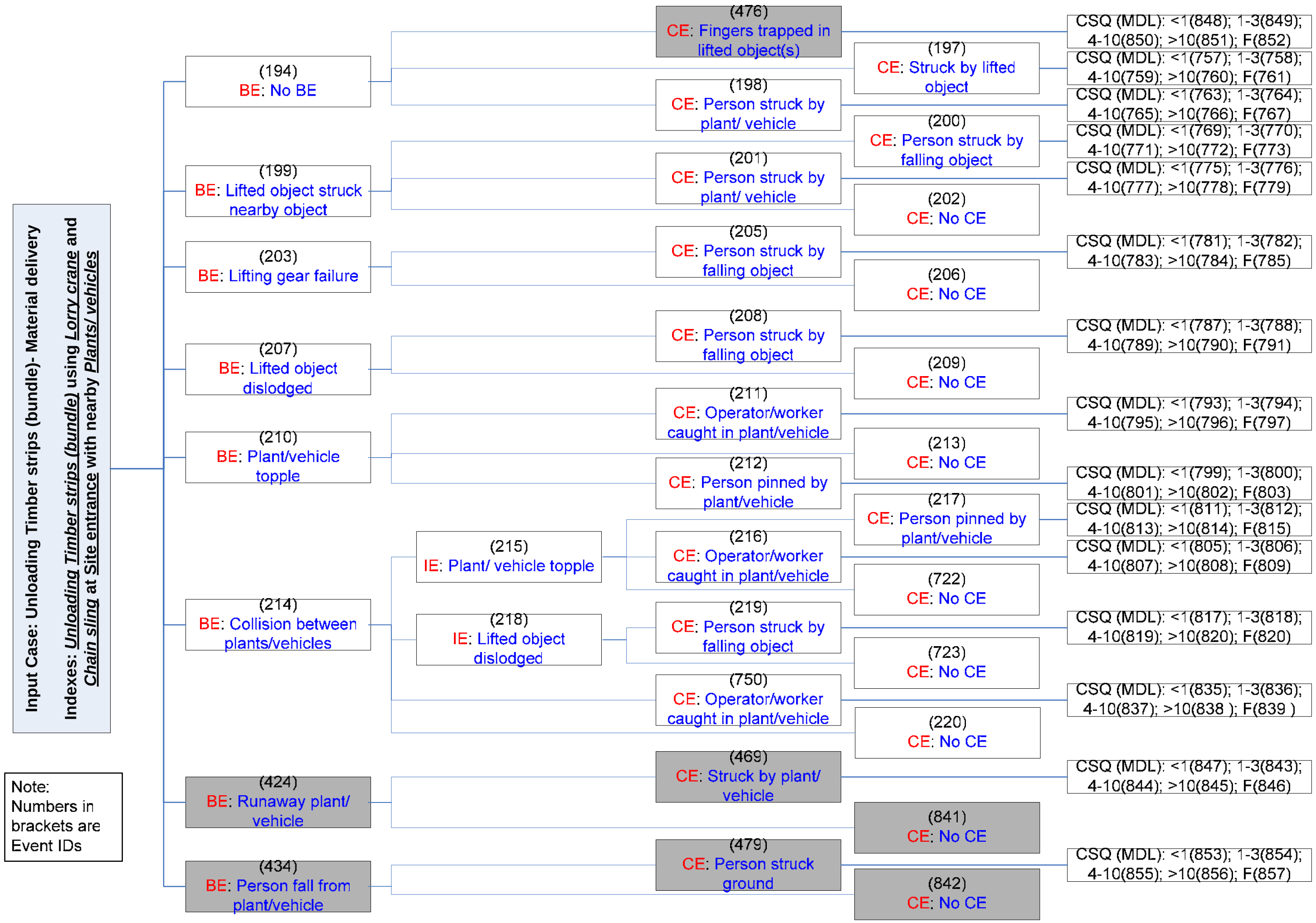


Figure 6.8 Risk assessment tree after hazard identification adaptation

data of cases belonging to projects with dissimilar SMS quality and/or project will have to be filtered off.

6.3.3.1 Adjustment of likelihood values to ensure consistency

Due to the removal and insertion of incident events during hazard identification adaptation, the likelihood values of the retrieved risk assessment tree will have to be adjusted to ensure consistency with Equations (5.8) and (5.9). The equations are based on the validated Poisson distribution, its assumption of statistical independence and the partitioned Poisson Model (refer to chapter 5). The following equations are replicates of Equations (5.8) and (5.9),

$$\overline{\lambda_i} = \sum_j \overline{\lambda_{ij}} \qquad (6.1)$$

$$Var(\lambda_i) = \sum_j Var(\lambda_{ij}) \qquad (6.2)$$

where $\overline{\lambda_i}$ is the mean frequency of process i measured in number of *incident per 50,000 mhr*, $\overline{\lambda_{ij}}$ is the mean frequency of sub-process *ij,* $Var(\lambda_i)$ is the variance of λ_i, and $Var(\lambda_{ij})$ is the variance of λ_{ij}.

With reference to Figure 6.9 (a blown up of Figure 6.6), under the BE, "No BE" [1](Event 194), there are three possible CEs, "Cut by object" (Event 195), "Struck by lifted object" (Event 197) and "Person struck by plant/ vehicle" (Event 198). When the CE, "Cut by object" (Event 194) was deleted, $\overline{\lambda_{194}}$ and $Var(\lambda_{194})$ have to be

[1] "No BE" refers to an incident that occurs with a direct contact event and no observable breakdown event.

adjusted so that $\overline{\lambda}_{194} = \overline{\lambda}_{197} + \overline{\lambda}_{198}$ and $Var(\lambda_{194})$ = $Var(\lambda_{197})$+ $Var(\lambda_{198})$, in accordance to Equation (6.1) and Equation (6.2). After the preceding incident event of the deleted incident event had been adjusted, the adjustment is then propagated towards the root node of the tree, thus ensuring the consistency for the whole tree.

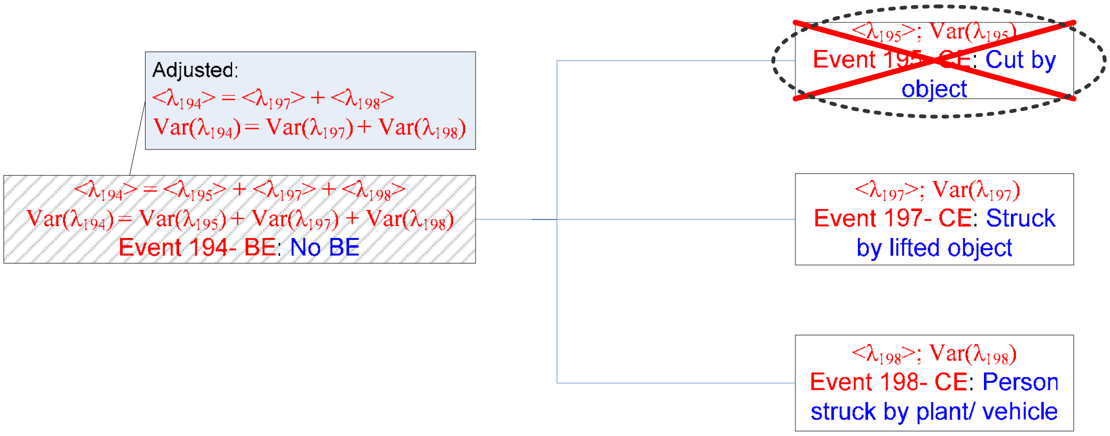


Figure 6.9 Adjustment of likelihood to account for deleted incident event

Besides adjustments for deleted incident events, inserted incident events will also require similar modifications. For example, with reference to Figure 6.10 (a blown up of Figure 6.8), the safety expert retained the prior subjective values of events 197 and 198, and estimated that the ratio of the likelihood values of Events 476 (inserted event), 197 and 198 to be 2 : 3 : 1. This yields a subjective estimate of 0.0041 *incidents per 50,000 mhr* for $\overline{\lambda}_{476}$, which is in-between $\overline{\lambda}_{197}$ = 0.0062 *incidents per 50,000 mhr* and $\overline{\lambda}_{198}$ = 0.0021 *incidents per 50,000 mhr*. The safety expert also assigned a coefficient of variation (C.O.V.) of 33.33% (or 1/3) to reflect his uncertainty and this gives $Var(\lambda_{476})$ = 1.88e-006. With the inserted Event 476, $\overline{\lambda}_{194} = 0.0062 + 0.0021 + 0.0041 = 0.012$ and $Var(\lambda_{194})$ = 4.76e-005 + 1.58e-005 + 1.88e-006 = 6.53e-005. The remaining events of the risk assessment tree shown in

Figure 6.8 are also treated similarly. The $\bar{\lambda}$ and $Var(\lambda)$ of the entire adjusted risk assessment tree can be found in Figures A5.1 to A5.9 of Appendix 5 and Tables A6.1 to A6.4 of Appendix 6.

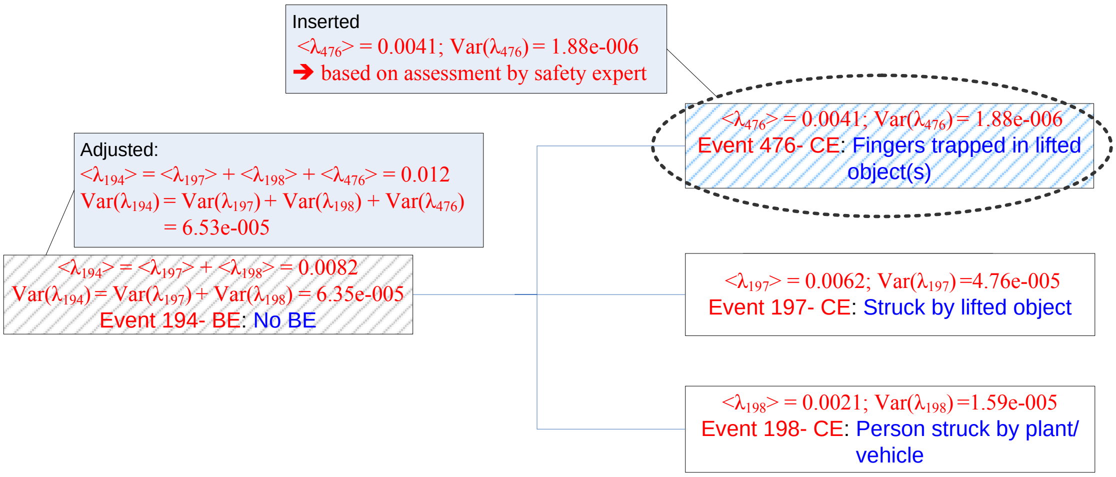


Figure 6.10 Adjustment of likelihood to account for inserted incident event

6.3.3.2 Bayesian updating

The next stage of the risk analysis is to update the prior subjective estimates of incident events in the finalised risk assessment tree by integrating the relevant retrieved data with the subjective estimates through the Bayesian approach. The Bayesian approach is based on the following set of equations, which were discussed in chapter 5 and replicated here for easy reference.

$$\bar{\lambda} = \kappa/\upsilon \qquad (6.3)$$

$$Var(\lambda) = \kappa/\upsilon^2 \qquad (6.4)$$

$$\kappa'' = \kappa' + x \qquad (6.5)$$

$$\upsilon'' = \upsilon' + t \quad (6.6)$$

where κ and υ are the parameters of the gamma-distributed λ, x refers to the number of actual incident occurring in t intervals (*50,000 mhr per interval*). It is noted that variables with an apostrophe ($'$) refers to prior estimates, whereas variables with a double apostrophe ($''$) refers to posterior or updated estimates.

In this case study, the case base contains incident cases of only one project. Thus, the t-value (Equation (6.6)) used in the Bayesian updating of the leaf nodes of the risk assessment tree will be the project's total *mhr*, which is 279 (50,000 *mhr* intervals). The leaf nodes refer to either consequence events (CSQ) or contact events (CE) of "No CE" type (see Figure 6.8). In terms of x-values (Equation (6.5)), out of the 84 leaf nodes, only five (Event ID = 206, 790, 844, 849 and 856) have one relevant incident case each (x = 1), and the rest have none (x = 0). Using Equations (6.3) to (6.6) and the calculation method presented in Section 5.3.3, the updated frequency values and variances of the leaf nodes are presented in Tables A6.1 to A6.3.

Subsequently, to ensure consistency with Equations (6.1) and (6.2), the updated frequency value and variance are propagated from the leaf nodes towards the root node of the risk assessment tree. During the propagation, the frequency values and variances of the parent nodes are updated to be equal to the sum of its child nodes' values. The propagation process is similar to the adjustment of likelihood values discussed in Section 6.3.3.1. The updated frequency values and variances of all the incident events are shown in Tables A6.1 to A6.5, while Table A6.6 of Appendix 6 contains the changes in the values of respective incident events.

If the severity categories can be assigned a severity value, risk values can be computed. In this case study, the severity categories are based on the number of man-days lost (MDL). As be seen in Figure 6.8, the severity categories are “<1 MDL”, “1-3 MDL”, “4-10 MDL”, “>10 MDL” and “Fatal (F)”. Based on consultations with safety experts, a quantified value is then assigned to each category to reflect the relative severity of each category (see Table 6.5). For non-fatal categories, i.e. “<1”, “1-3”, “4-10” and “>10”, the quantified value is based on the expected MDL of the category. For the fatal category, the value of 100 was assigned to reflect the high severity of an incident with fatality, but at the same time not cause risk values to be overly-sensitive to frequency estimates of fatal categories. It was noted that Singapore’s Ministry of Manpower (MOM) assigns an arbitrary value of 6000 MDL for each fatal incident for their calculation of incident statistics, but the large value might cause bias in the calculation of risk values and was not used in this research. The severity categories and values may be modified to reflect an organisation’s perception of the relative impact of the various severity categories.

Table 6.5 Quantified severity value for severity categories used in the case study

Severity Category	Quantified Severity
< 1 MDL	0.5
1-3 MDL	2
4-10 MDL	7
>10 MDL	20
Fatal	100

The updated risk assessment tree can then be used to determine the overall risk of the job step and the individual risk of different incident events. That is,

$$Risk(SV) = \overline{\lambda} \bullet \sum_i P(B_i) \bullet \sum_j P(I_{ij}) \bullet \sum_k P(C_{ijk})[\sum_l S_{ijkl} \bullet P(S_{ijkl})] \qquad \textbf{(6.7)}$$

where SV refers to the set of situational variables that represents the job step, $\overline{\lambda}$ is the mean incident occurrence rate of the job step, B_i refers to breakdown event i, I_{ij} refers to intermediate event j under B_i, C_{ijk} refers to contact event k under I_{ij}, and S_{ijkl} refers to severity category l under C_{ijk}.

Equation (6.7) can also be interpreted as,

$$Risk(SV) = \sum_i \sum_j \sum_k \sum_l (\overline{\lambda} \bullet P(B_i) \bullet P(I_{ij}) \bullet P(C_{ijk}) \bullet P(S_{ijkl}) \bullet S_{ijkl}) \qquad \textbf{(6.8)}$$

where $[\overline{\lambda} \bullet P(B_i) \bullet P(I_{ij}) \bullet P(C_{ijk}) \bullet P(S_{ijkl})]$ refers to the expected frequency of S_{ijkl}, or $\overline{\lambda}_{ijkl}$. In this way, the risk of the job step can be determined based on the following equation,

$$Risk(SV) = \sum_i \sum_j \sum_k \sum_l (\overline{\lambda}_{ijkl} \bullet S_{ijkl}) \qquad \textbf{(6.9)}$$

Thus, the risk of the job step can be determined by simply summing up the risk values ($\overline{\lambda}_{ijkl} \bullet S_{ijkl}$) of all the consequence events in the risk assessment tree. This is advantageous because the SKMS keeps track of the expected frequency of incident events and not the probability values.

Likewise, the risk of different incident events can be determined by summing up the risk values of all the consequence events under it. For instance, the risk of breakdown event i,

$$Risk(B_i) = \sum_j \sum_k \sum_l (\overline{\lambda}_{ijkl} \bullet S_{ijkl}) \qquad \textbf{(6.10)}$$

Based on the frequency estimates and severity values, the risk values of all the incident events and the overall risk of the job step was determined and recorded in Tables A6.1 to A6.5 of Appendix 6. For easy reference and discussion, Table 6.6 shows the prior and posterior frequency, variance, severity and risk values of the overall job step and breakdown events. Table 6.7 shows the change (posterior – prior) in the various values after the risk analysis adaptation.

With reference to Table 6.6, the job step's risk value before and after Bayesian update is 0.0525 and 0.0293 respectively. Table 6.7 shows the effect of the Bayesian updating in more detail. As can be seen, all the frequency values and variances decreased. The drop in frequency is due to the lowered frequency values of most of the leaf nodes (refer to Tables A6.1 to A6.3 of Appendix 6), which decreased because most of the objective data has a lower frequency than the prior estimates. The reduction in variance reflects a higher certainty on the estimation of the various $\overline{\lambda}$ due to the integration of objective data into prior subjective estimation.

In terms of severity, four of the breakdown events (BEs) have decreased severity (Event ID = 194, 203, 207, and 424), one has increased severity (Event ID = 434) and three have no change (Event ID = 199, 210, and 214). The severity of Events 199, 210 and 214 did not change because all the leaf nodes under these BEs were updated with the same set of data (x = 0 and t = 278.966) and all the frequency value changed proportionately. This maintained the distribution of the various severity categories and hence the expected severity is unchanged. On the other hand, Event 194, 203, 207, 424 and 434 each has one leaf node updated with the data, x = 1 and t = 278.966. In this way, the distribution of the severity category is altered and hence it results in changes to the expected severity. It is noted that only Event 434 has

Table 6.6 Prior and posterior frequency, variance, severity and risk values of overall job step and breakdown events

ID	Event Type	Event Value	Prior				Posterior			
			$\bar{\lambda}'$	Var(λ)	*E*(Sev)	Risk	$\bar{\lambda}''$	Var(λ)	*E*(Sev)	Risk
840	Root	Root	0.0525	3.49E-04	26.7	1.40	0.0293	5.57E-05	21.4	0.626
194	BE	No BE	0.0123	6.53E-05	16.9	0.209	0.00666	8.03E-06	14.7	0.0982
199	BE	Lifted object struck nearby object	0.0055	4.23E-05	35.3	0.194	0.00174	4.26E-06	35.3	0.0614
203	BE	Lifting gear failure	0.0128	9.87E-05	33.6	0.430	0.00651	1.59E-05	20.9	0.137
207	BE	Lifted object dislodged	0.00914	7.05E-05	33.6	0.307	0.00535	1.31E-05	27.4	0.146
210	BE	Plant/vehicle topple	0.00183	1.41E-05	23.5	0.0429	0.00058	1.42E-06	23.5	0.0136
214	BE	Collision between plants/vehicles	0.00365	2.82E-05	24.8	0.0906	0.00116	2.84E-06	24.8	0.0287
424	BE	Runaway plant/ vehicle	0.00365	1.48E-06	23.1	0.0843	0.00365	1.33E-06	21.5	0.0783
434	BE	Person fall from plant/vehicle	0.00365	2.82E-05	12.0	0.0440	0.00361	8.83E-06	17.4	0.0629

Table 6.7 Difference between prior and posterior frequency, variance, severity and risk values of overall job step and breakdown events

Event ID	Event Type	Event Value	Difference (Posterior – Prior Estimates)			
			$\bar{\lambda}$	Var(λ)	*E*(Sev)	Risk
840	Root	Root	-0.0233	-2.93E-04	-5.27	-0.776
194	BE	No BE	-0.00567	-5.73E-05	-2.18	-0.111
199	BE	Lifted object struck nearby object	-0.00374	-3.80E-05	0	-0.132
203	BE	Lifting gear failure	-0.00629	-8.28E-05	-12.7	-0.294
207	BE	Lifted object dislodged	-0.00379	-5.74E-05	-6.24	-0.161
210	BE	Plant/vehicle topple	-0.00125	-1.27E-05	0	-0.0293
214	BE	Collision between plants/vehicles	-0.0025	-2.54E-05	0	-0.0618
424	BE	Runaway plant/ vehicle	-7.2E-06	-1.54E-07	-1.61	-0.00603
434	BE	Person fall from plant/vehicle	-4.8E-05	-1.94E-05	5.40	0.0189

increased severity. This is due to increased frequency value for leaf node 856 (see Table A6.3 of Appendix 6), which has a severity category, “>10” (severity value = 20) (refer to Table 6.5), that is higher than the initial expected severity of Event 434 (E(Sev) = 12.05) (see Table 6.6). The higher frequency of leaf node 856 causes the expected severity to be swayed towards the severity value of 20. The adjustment is expected because the knowledge of a high severity incident resulting from a BE will logically cause the expected severity to be adjusted upward. For Events 194, 203, 207, and 424, the expected severity decreased because the data x = 1 and t = 278.966 was updated on a leaf node that has a lower severity value than its initial severity value.

Table 6.7 also shows that the overall risk values of all the BEs have decreased and hence the risk of the job step has also decreased. This is expected because the severity and frequency of all the BEs, with the exception of Event 434’s severity value, have been updated to a lower value. In the case of Event 434, the risk value has increased slightly because the increase in the severity value (+44.8%) is more than the decrease in the frequency value (-1.1%)

To clearly identify and evaluate the risk levels of various job steps and incident events, a risk contour plot can be created. A risk contour plot is usually based on a risk matrix, such as the one shown in Table 6.8, which was developed based on the severity and frequency categories of Tables 6.5 and 6.9 respectively. The frequency categories of Table 6.9 were based on the categories used by LTA and the $\overline{\lambda}$ values were determined using the average *mhr* per month for LTA projects captured in the SITS between 1998 and 2002. The set of $\overline{\lambda}$ values were also verified with safety experts. Table 6.8 also shows the acceptability levels of various risk values assigned by one of the safety experts who participated in the case study. The risk contour of Figure 6.11 is developed based on the risk matrix of Table 6.8.

Table 6.8 A risk matrix developed for the case study

Severity	Likelihood				
	Improbable (0.0127)	Remote (0.0254)	Occasional (0.0762)	Probable (0.152)	Frequent (0.915)
Fatal (100)	1.270	2.54	7.62	15.2	91.5
> 10 MDL (20)	0.254	0.508	1.52	3.05	18.3
4-10 MDL (7)	**0.0889**	**0.178**	0.534	1.07	6.40
1-3 MDL (2)	*0.0254*	**0.0508**	**0.152**	0.305	1.83
< 1 MDL (0.5)	*0.00635*	*0.0127*	**0.0381**	**0.0762**	0.457

Unacceptable - Underlined   Undesirable - Shaded
Acceptable - Italics   **Tolerable - Bold**

Table 6.9 Quantified likelihood (λ) value for various frequency categories

Likelihood	Definition (Expected frequency)	λ (Incident per 50,000 *mhr*)
Frequent	1 incident monthly	0.915
Probable	1 incident 6 monthly	0.152
Occasional	1 incident yearly	0.0762
Remote	1 incident 3 yearly	0.0254
Improbable	1 incident 6 yearly	0.0127

Note: The λ-values are calculated based on 54,665 *mhr* per month, which is the average *mhr* per month for LTA projects captured in the SITS (1998-2002)

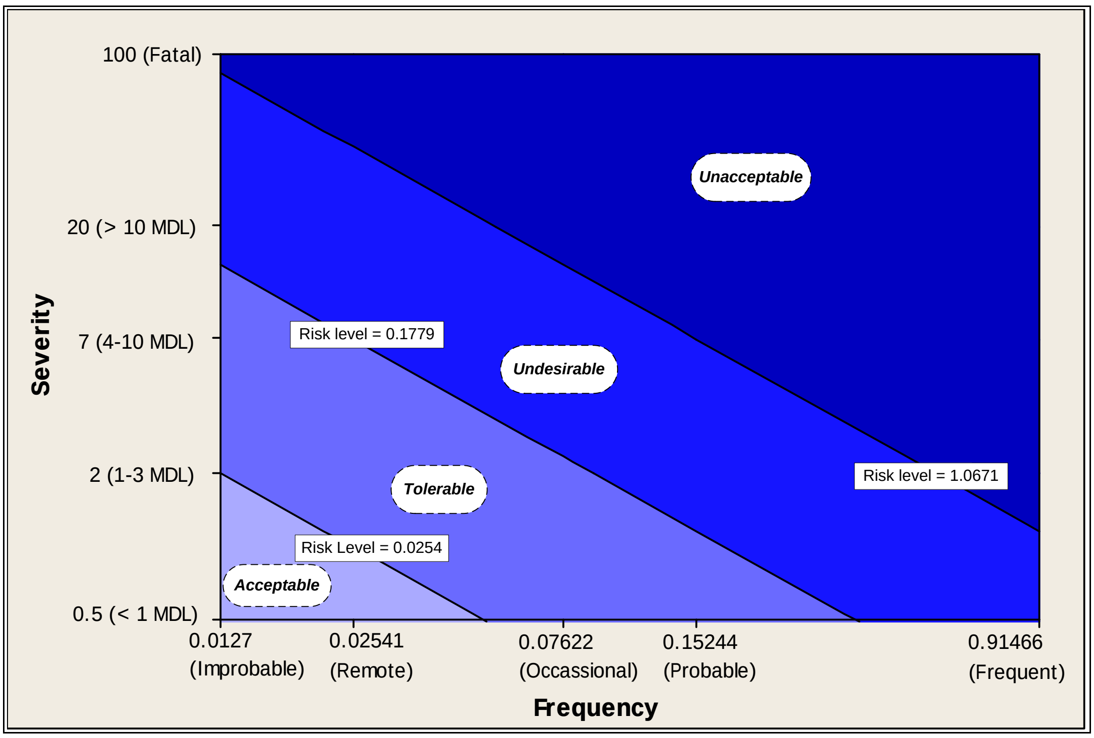


Figure 6.11 A risk contour plot indicating different levels of risk acceptability

To demarcate the acceptability of different risk values, the highest risk value in the respective expert-assigned acceptability categories are used to determine the boundaries (see Table 6.8). This is to ensure that the risk values between the lowest risk value of a less acceptable category and the highest risk value of a more acceptable is classified as the less acceptable category. For instance, with reference to Table 6.8, the highest risk value in the "Undesirable" category is 1.07 ("Probable" and "4-10 MDL") and the lowest risk value in the "Unacceptable" category is 1.23. Based on the convention adopted, the risk values between these two values will be conservatively classified as "Unacceptable" and the boundary between the "Unacceptable" and "Undesirable" categories is demarcated by the risk value 1.07. The same approach is also used to determine the other boundaries.

The apriori and aposteriori risk values of the job step and incident events derived from the risk analysis can then be plotted on the risk contour plot (see Figure 6.12) for further analysis. In this way, the activities with unacceptable and undesirable risk levels can be easily identified and prioritised for risk control selection and implementation.

6.4 Discussions

Generally, the case study had demonstrated that the prototype SKMS is able to facilitate feedback of safety knowledge through the proposed case representation scheme, retrieval mechanism and adaptation strategies. The resulting risk assessment tree is a combination of the knowledge gleaned from past experiences and the user's input.

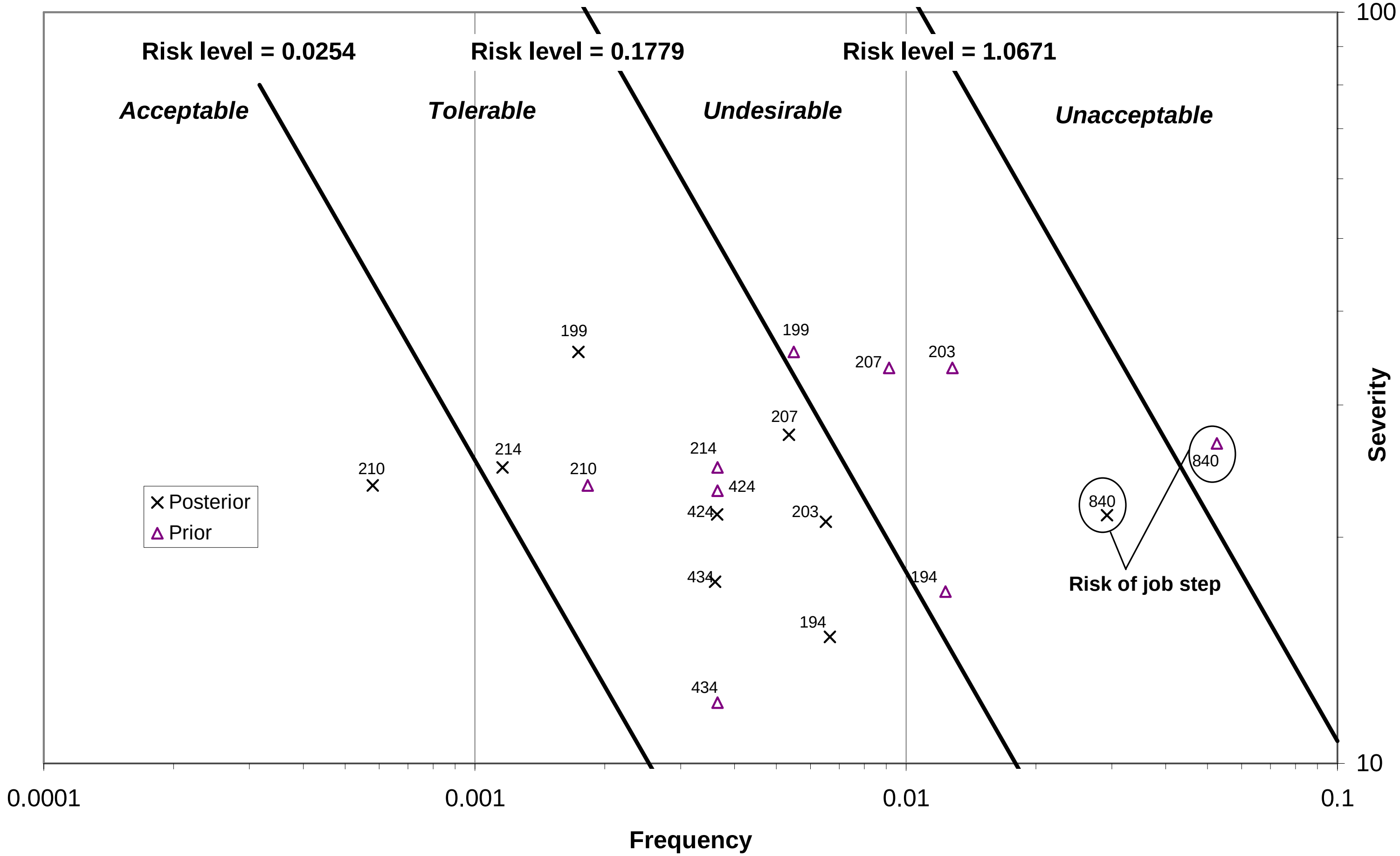


Figure 6.12 Risk contour plot with prior and posterior risk values of job step and incident events

6.4.1 Retrieval

The proposed retrieval method utilises the knowledge stored in the semantic networks and the weights captured in past cases to measure similarity between stored cases and the input case. The LSS and GSS are then used to identify and retrieve relevant stored cases. Both LSS and GSS are important information to the users because they provide specific information on why a stored case is relevant. The user can easily understand the rationale of the retrieval by assessing the LSS and GSS of each case. The transparent approach also allows users to evaluate the appropriateness of the retrieval and make adjustments whenever necessary.

As compared to manual search, or database keyword search, the proposed retrieval mechanism has obvious advantages. Most importantly, the similarity scoring approach allows similar (but not exactly matching) cases to be retrieved and utilised. It is evident from Table 6.4 that all the retrieved incident cases do not have exactly matching indexes as the input case (GSS < 1). Thus, if an exact matching approach is used these similar cases would not have been retrieved and important information on hazards and risk would be lost.

The case study also showed that users only need to describe the situational variables of the input case and relevant cases can be retrieved. This characteristic allows less experienced personnel to be able to produce detailed risk assessment trees with more convincing risk values. Furthermore, the threshold similarity score used to demarcate relevant cases is adjustable. Therefore, users can always vary the threshold to reach a balance between relevance and number of retrieved cases.

6.4.2 Hazard Identification

The similarity scoring approach simulates human cognition by retrieving similar cases for reuse, but these past cases may contain irrelevant portions that should be trimmed away. Thus, as demonstrated in the case study, the risk assessment tree is pruned to remove incident events deemed to be irrelevant. The rationale for the adaptation is transparent and can be modified by the user when necessary. For example, in the case study the contact event, “Cut by object”, was deleted because the index of the incident event, “Object-worked-on (OWO) = Pipe” has a low LSS with the input case’s situational variable, “OWO = Timber strip (bundle)”. The user may still want to include the incident event because he might regard it as a possible incident event even if the OWO is a bundle of timber strips. Since each adaptation is based on the LSS, the user can easily review and verify the basis of hazard identification adaptation.

Besides pruning, the hazard identification adaptation also includes integration or insertion of incident cases into the retrieved risk assessment tree. The adaptation strategy ensures that relevant incidents cases that are not already in the risk assessment tree are inserted. In this way, risk assessment teams will always be reminded of relevant incident occurrences, allowing measures to be implemented to prevent recurrence.

6.4.3 Risk Analysis

The SKMS risk analysis approach first ensures that the likelihood values in the retrieved risk assessment tree is consistent throughout the tree, and then the prior likelihood estimates are integrated systematically with objective estimates from incident cases. This approach ensures that the prior estimates are balanced with objective data that would have been considered statistically insignificant if used on its own. The Bayesian approach also facilitates continual learning because it allows the likelihood estimates to become more objective progressively through integration of objective data at each update.

In this case study, the job step's risk value before Bayesian update is 1.402 and after the Bayesian update the risk value is 0.626 (refer to Table 6.6). As can be observed from the risk contour plot of Figure 6.12, the prior risk value of the job step was in the "Unacceptable" region, but after the Bayesian updating, the risk level was adjusted to be in the "Undesirable" region. This reduction shows that the prior frequency and severity estimates might be on the high side.

As presented in Section 6.3.3.2, the subjective prior frequency estimate of 0.0525 is reduced because the objective data (x = 5 and t = 279) gives a lower frequency estimate of 0.0179 (x/t). Thus, when the subjective and objective estimates are integrated the posterior frequency is 0.02925, which is in-between the subjective and objective estimates. Similarly, as noted in 6.3.3.2, the expected severity of the job step is reduced because most of the retrieved incident has a lower severity value than the prior expected severity of corresponding breakdown events. This was verified with the safety expert who had given the prior estimates. He agreed that his estimation tends to be more conservative which could have led to higher prior values. Hence, the case study has shown that the Bayesian approach was able to systematically balance the conservative estimation with actual observations obtained from the incident cases.

The main purpose of the risk contour plot of Figure 6.12 is to allow the risk assessment team to easily identify high risk job steps and spend more resources to reduce the risk levels of these job steps. The risk contour plot can also facilitate evaluation of the change in frequency, severity and risk of the overall job step and the various incident events. Risk assessment teams can easily determine the reasons for changes in risk values. For instance, a horizontal shift in the risk values on the plot indicate that the change in risk is due to a change in frequency only. Besides that, the contour plot also allows the risk assessment team to quickly identify high risk BEs to focus on to decrease the risk level of the job step. Furthermore, if the risk assessment team chooses to focus on high severity events (e.g. Events 199 and 207) in the same risk category, these events can also be easily identified through the plot.

In the case study, all the stored cases can be used for the Bayesian updating because the SMS quality and project type is similar to that of the input case. This is necessary to ensure consistency with the proposed Poisson model (see Figure 5.11). This constraint can be easily overcome as the case base grows, because in any learning process the initial stage is usually tougher, but once more knowledge is gained, the system will be able to deal with a larger variety of situations.

6.5 Conclusions

The case study had demonstrated how the proposed methodologies and concepts will work in a realistic scenario. The case study also showed how the calculated risk assessment tree and risk values can be utilised to facilitate prioritising of risk control efforts. One of the key advantages of the SKMS is that users can easily understand its retrieval and adaptation decisions and modify whenever necessary.

The SKMS will be able to facilitate feedback of past safety experiences to improve the current risk assessment process. The feedback helps users to identify possible hazards, prevent recurrence of past incidents, provide a basis for likelihood assessment and improve efficiency of the risk assessment process. Such feedback is important in ensuring continual improvement and organisational learning.

Chapter 7

CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

The basic purpose of this research is to encourage continual improvement of construction projects' safety management systems (SMS) through learning from past experiences. To enable construction companies to learn from safety knowledge accumulated in the form of incident cases and past safety plans, this research developed a novel case-based reasoning (CBR) approach to safety planning. The proposed approach models and stores safety knowledge in the appropriate knowledge framework, and fully utilises them during safety planning through unique retrieval and adaptation strategies. This research had been focused on feedback of safety knowledge to the risk assessment component (hazard identification and risk analysis) of a safety planning process. However, the principles and methodologies developed are also applicable to the risk control selection component.

To achieve the desired feedback of safety knowledge during risk assessment, the following research components had been developed.

1. The Modified Loss Causation Model (MLCM) is meant to cover a wider scope than facilitating the proposed CBR approach to risk assessment. The MLCM is able to facilitate thorough incident investigation and hence facilitate the feedback of safety knowledge to improve the SMS that had failed and caused the incident. This provides the first level feedback. More importantly, the MLCM acts as a common knowledge structure for both incident investigation and safety planning. In this way, safety

knowledge can then be utilised to improve safety planning for new projects. This facilitates the second level feedback which will benefit safety planning across projects.

2. To implement the broad structure provided by the MLCM in the prototype CBR system, known as the Safety Knowledge Management System (SKMS), a detailed knowledge representation scheme was developed to abstract and capture safety knowledge in incident cases and past risk assessments. The knowledge representation scheme was necessary because the CBR system requires a specific framework to implement the retrieval and adaptation processes. The monolithic case approach, as opposed to a snippets approach, was adopted to facilitate the development of the knowledge representation scheme. The monolithic case approach is advantageous because cases are kept intact and not separated into sub-cases or snippets as in the latter approach. In this way, subtle details within a complete case which could have been missed if the case has been separated into snippets will be made available to the human user. Besides, the monolithic case approach is also computationally less expensive than the snippets approach. However, it was acknowledged that a snippets approach is fundamentally similar to the monolithic approach and can also be applied. The knowledge representation scheme was designed to facilitate the Job Hazard Analysis (JHA) process, where the situational variables or indexing vocabulary describe key parts of a job step. For each stored case, suitable situational variables are chosen to act as the indexes, which are necessary prerequisites for the case retrieval process. The situational variables also represent the possible hazards in the work

scenario and the types of situational variables include: “Action”, “Object-worked-on”, “Resource”, “Location”, “Nearby object” and “Nearby action”.

3. Through the knowledge representation scheme, an intelligent retrieval method that can automatically identify and retrieve relevant cases was created. The retrieval method is based on customised local and global similarity scoring functions that suits the context of the SKMS. The local similarity score (LSS) between two situational variables is calculated using the degree of match of weighted sub-concepts. The sub-concepts are important concepts related to hazardous objects, energies or harmful substances that have implications on the safety or risk of a job step. To facilitate the determination of weights of the sub-concepts, semantic networks of sub-concepts had been developed for different situational variables. The sub-concepts nearer to the root node of the semantic network (i.e. more general) or more directly related to potential hazards are assigned higher weights. The global similarity score (GSS) of the input case and each stored case is calculated using a weighting function applied on the LSS of relevant situational variables. Unlike conventional GSS functions, the weights are assigned when the cases are stored into the case base, instead of during retrieval. In the context of the SKMS, the weights assigned to the various situation variables reflect the importance of each variable in relation to the specific case. In this way, the GSS computed will be able to better reflect the relevance of each stored case during retrieval.
4. Adaptation strategies had also been developed to contextualise the retrieved cases for hazard identification and risk analysis. During hazard identification, adaptation strategy is meant to improve the relevance of the risk assessment tree by removing

irrelevant incident events of the retrieved cases. Furthermore, the adaptation mechanism also integrates all retrieved cases into one single risk assessment tree. The integrated risk assessment tree containing all the identified incident sequences is then utilised for risk analysis purposes. The adaptation strategy during risk analysis utilises the Bayesian approach, such that subjective and objective estimates of likelihood are integrated to provide a more realistic likelihood values. The Bayesian approach also facilitates continual improvement because subjective prior estimates can be improved whenever new incident cases are incorporated into the prior estimates. The reviewed likelihood values are then used to determine risk values for different job steps and incident events.

The abovementioned research components had been implemented in a prototype SKMS that is applied on a case study based on a typical construction work scenario. The case study utilises a case base with 59 actual incident cases from a single past project and 10 risk assessment trees obtained from safety experts, the Land Transport Authority (LTA) of Singapore, various main contractors and Mine Safety and Health Administration (MSHA 2004) of the United States. The results of the case study demonstrated that the proposed CBR approach to risk assessment can produce a reasonably thorough and well-balanced risk assessment tree through feedback of safety knowledge despite the relatively small size of the case base. This shows that the approach is able to capitalise on available knowledge and maximise their benefits. More project data, can subsequently be integrated and will improve the risk assessment.

The resulting risk assessment tree includes possible incident sequences identified by previous risk assessment team and actual incident occurrences contained in relevant incident cases. During retrieval, the most relevant risk assessment tree (GSS = 0.61) and five incident cases (GSS $\geq$ 0.6) were retrieved. The risk assessment tree was pruned to remove seven incident events that were considered irrelevant to the input case. Furthermore, the five incident cases also helped to identify new incident sequences, where five previously unidentified incident events were inserted. This shows that it is possible for safety and risk assessment teams to miss out certain incident sequences, especially when there are tight time constraints. Another five incident events of the retrieved incident cases duplicated incident events already identified in the risk assessment tree. These duplicated incident events contain important frequency data that are utilised during risk analysis. Thus the risk assessment tree is effectively based on reused or fed back safety knowledge and it successfully highlighted actual past occurrences to prevent recurrence.

Furthermore, the estimated risk values of the case study are also determined based on the integration of both retrieved risk assessment tree (subjective source) and incident cases (objective source). The prior estimated risk value of the job step was 1.402, and after the risk analysis adaptation, the risk value was lowered to 0.626. The risk acceptability level was thus amended from “Unacceptable” to “Undesirable” level. The change in risk level is a result of the integration of knowledge gleaned from both objective and subjective sources. In the case study, the subjective source tends towards the conservative end, while the objective source is only based on a single project. On its own both might not be able to provide a balanced estimate. Hence, by integrating both estimates, the Bayesian update provided a more realistic estimate of the risk of job steps and its incident events.

The case study also demonstrated how a risk contour plot can be utilised to aid prioritisation of risk control selection and implementation efforts. The contour plot, which is based on a risk matrix developed with the help of safety experts, allows ra isk assessment team to quickly recognise the risk acceptability of different activities, job steps, and incident events. Furthermore, severity and frequency values can also be quickly evaluated to allow the risk assessment team to understand the nature of the risk. The risk contour plot also facilitates understanding of the impact of Bayesian updating. The risk assessment team can observe the relative positions of the prior and posterior risk values on the plot and easily deduce the change in severity and frequency values that caused the change in the risk levels. This understanding allows further modifications based on the judgement of the risk assessment team.

7.2 Limitations and Recommendations

This research had spearheaded a worthwhile direction in construction safety and developed a novel CBR approach for risk assessment, but more can still be done to build on the foundations provided by this research to further improve effective feedback of safety knowledge. The following limitations and/or recommendations are noted and discussed.

1. In this research, only the most relevant risk assessment tree is retrieved to facilitate risk assessment. However, to ensure that more possible incident sequences are identified, it may be desirable to utilise all relevant risk assessment trees. This would mean that the risk assessment trees will have to be integrated into a single risk assessment tree. The integration of risk assessment trees is similar to the insertion of incident cases into the most similar risk assessment tree, but this operation has a

relatively high computational cost. That is because the number of comparison between incident events belonging to different risk assessments can be very large. Future research can explore more efficient indexing algorithms to reduce the computational cost involved in the integration process.

2. The SKMS is a learning system that needs to accumulate more cases to ensure that it is able to provide relevant knowledge. Nevertheless, the case study demonstrated that a "young" SKMS is able to capitalise on small amount of available knowledge and provide reasonably thorough assessment of the hazards and risk in a realistic scenario. Still, to ensure that the SKMS can cover a wide array of work situations, consistent efforts will be needed to build, codify and store cases into the case base. These efforts may be minimised by integrating the knowledge representation efforts into existing safety planning and incident investigation processes, so that safety knowledge can be directly captured from the risk assessment tree and incident cases. These efforts are definitely worthwhile because they will significantly increase the utility of the stored safety knowledge and prevent the phenomenon of "data graveyard" where large amount of data is stored but serve no substantial purposes.
3. One of the key assumptions of the partitioned property of the Poisson model is that the categorisation of an incident is random. As indicated in chapter 5, this assumption is reasonable because the set of situational variables are low level descriptions of construction work, such that it is generally applicable to most possible types of site activities. However, in the event that a specialised type of work is executed within a specific period of the project, the actual start and end date of the activity need to be recorded to allow accurate estimation of incident frequency or likelihood values. If

the actual work period is not specified, the likelihood estimation during risk analysis would be deflated erroneously due to the utilisation of the whole project timeline instead of the shorter work period of the specialised activity. Once the correct work period is identified, the rest of the computation for the adaptation of likelihood value would be similar to the proposed methods discussed in chapters 5 and 6. Furthermore, it is noted that the Poisson model is only a rough approximation of construction incident occurrence and it may be problematic when representing projects near deadlines, with tight budgets and schedules. It is recommended that more data to be collected to further validate the Poisson model. Other more sophisticated distributions should also be examined and tested for their ability to model a wider variety of construction projects, especially those with significant resource constraints.

4. In a well-managed project there should be proper and consistent reporting of incidents and the problem of non-reporting should be minimal. This is especially so for incidents with higher severity, because these incidents are usually legally required to be reported to the authorities (such as severity > 3 MDL in the case of Singapore). In the situation where non-reporting of incidents is of concern, the concept of partitioned Poisson can also be used to account for unreported incidents. This is achieved by assuming that incidents are randomly partitioned into unreported and reported incidents. If the probability of an incident being reported is p_r then the true distribution of incident occurrence can be estimated by a Poisson distribution with parameter $\overline{\lambda}/p_r$, where $\overline{\lambda}$ is the estimated mean occurrence rate of reported incidents. $\overline{\lambda}$ can be obtained based on the risk analysis process presented earlier. The estimation of p_r can be based on expert opinion or statistical studies similar to the

study by Alsop and Langley (2001) on traffic incidents. Another possible approach is to assign a value for p_r based on the situational variables and each incident sequence in the risk assessment tree (refer to Figure 7.1). Thus,

$$p_r = P(r|SV, B_i, C_j, S_k) \tag{7.1}$$

where r is the event when an incident is reported, SV the situational variables, B_i the breakdown event i, C_j the contact event $_j$, and S_k the severity category j. In this way, the risk assessment team will assign a specific p_r for each branch of the risk assessment tree. These p_r refer to the probability values at the bottom of the risk assessment tree in Figure 7.1. The risk analysis adaptation is then applied on the r (reported) event of the branches instead of the consequence events to obtain the frequency of the r event, $\overline{\lambda_r}$. The frequency of each consequence event is then estimated based on $\overline{\lambda_r}$ $/p_r$. Subsequently the updated frequency values are then propagated towards the root node as discussed in chapters 5 and 6.

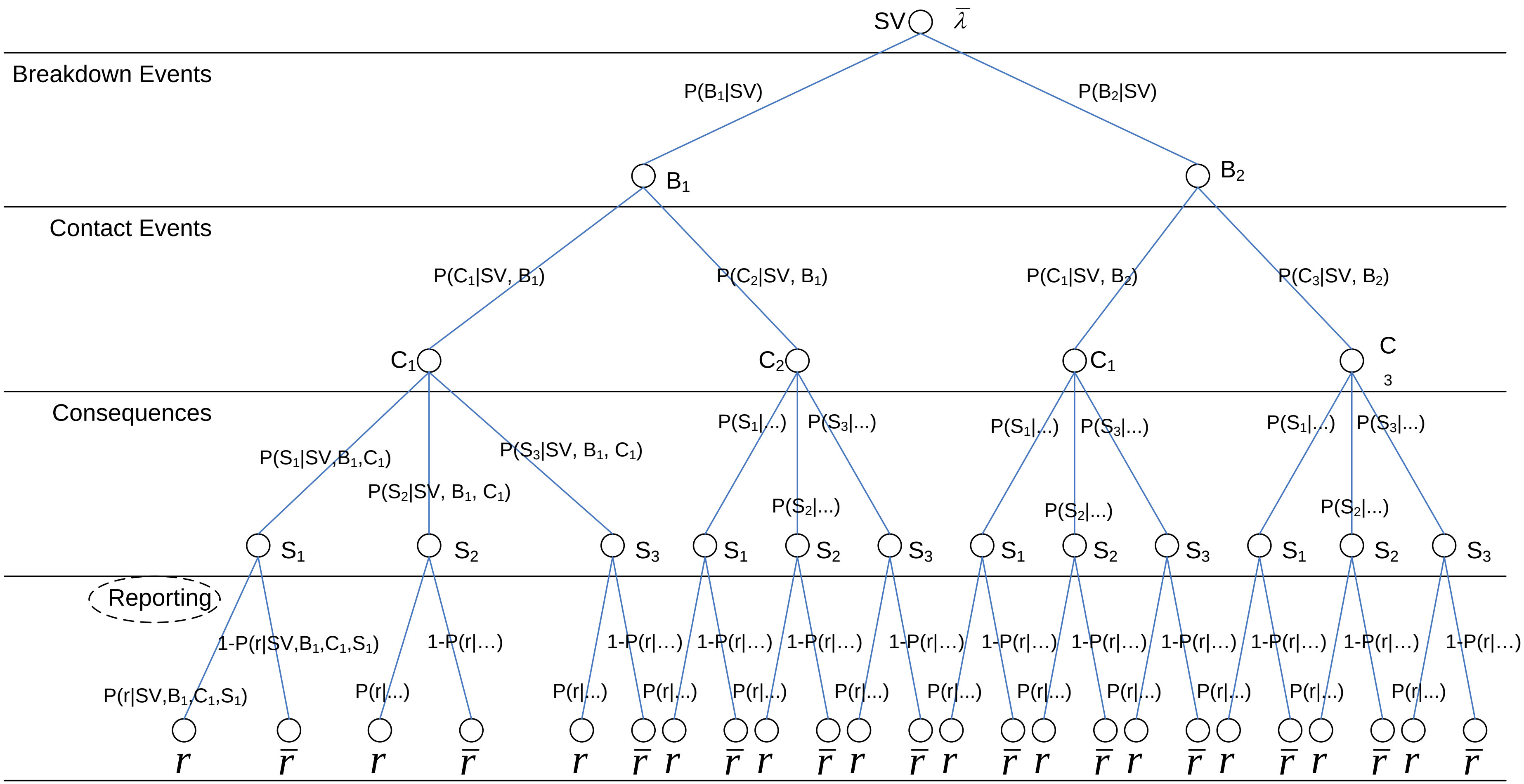


Figure 7.1 Risk assessment tree to account for non-reporting of incidents

REFERENCES

Abdelhamid, T. S. and Everett, J. G. (2000). "Identifying root causes of construction accidents." *ASCE, J. Constr. Engrg. and Mgmt.*, 126(1), 52-60.

Alsop, J. and Langley, J. (2001). "Under-reporting of motor vehicle traffic crash victims in New Zealand". *Accident analysis and prevention*, 33, 353-359.

Ang, H. S. and Tang, W. H. (1975). *Probability concepts in engineering planning and design Vol. 1- Basic Principles*. John Wiley Inc., New York.

Antelman, G. (1997). *Elementary Bayesian statistics*. Edward Elgar Publishing, New York.

Ashley, K.D. (1988). "Arguing by analogy in law: A case-based model". In *Analogical reasoning: Perspectives of artificial intelligence, cognitive science, and philosophy,* Helman, D. (ed), Kulwer, Boston.

Ayyub, B. M. (2003). *Risk analysis in engineering and economics.* Chapman and Hall/CRC, London.

Baets, W. R. J. (1998). *Organisational learnings and knowledge technologies in a dynamic environment*. Kluwer Academic Publishers, The Netherlands.

Baker, M. J., Graves, R. J. and Kearney (1996). "A methodology for risk management in civil, mechanical and structural engineering", In Proc. *Risk Management in Civil, Mechanical and Structural Engineering*. Thomas Telford, London.

Bendell, A. (1991). "Analysis methodologies". *Reliability data banks*. Cannon, A. G., and Bendell, A. (eds). Galliard Ltd, Great Britain.

Bergmann, R., Wilke, W. (1998). "Towards a new formal model of transformational adaptation in case-based reasoning", *In Proc. European Conference on Artificial Intelligence 1998 (ECAI '98).* Available at: http://www.agr.informatik.uni-kl.de/Bbergmann/Papers/ecai98.ps.gz.

Bird, F.E. and Germain, G.L. (1996). *Practical Loss Control Leadership.* Det Norske Veritas, Georgia.

Börner, K. (1995). *Modules for design support,* FABEL-Report No. 35, Centre for Cognitive Science. University of Freiburg, Germany.

Brown, J. S. and Duguid, P. (1998). "Orgainsing knowledge", *California Management Review*, 40(3), 90-112.

BSI. (1996). "Guide to occupational health and safety management systems", *BS 8800*. BSI, London.

BSI. (2000). "Occupational health and safety management systems- Guidelines for the implementation of OHSAS 18001" *OHSAS 18002*. BSI, London.

Catino, C.A. and Ungar, L. H. (1995). "Model based approach to automated hazard identification of chemicals", *AiChe Journal*, 41(1), 97109.

Cattledge, G. H., Hendricks, S. and Stanevich, R. (1996). "Fatal occupational falls in the U.S. construction industry, 1980-1989", *Accident Analysis & Prevention*, 28(5), 647-654.

Chae, H., and Yoon, Y. H. (1994). "Safety analysis using expert system in chemical processes", *Korean Journal of Chemical Engineering*, 11(3), 153-161.

Chua, D. K. H., Kog, Y. C., Loh, P. K., and Jaselskis, E. J. (1997). "Model for construction budget performance- neural network approach", *ASCE, J. Constr. Engrg. and Mgmt.*, 123(3), 214-222.

Chua, D. K. H., Li, D. Z. and Chan, W. T. (2001). "Case-based reasoning approach in bid decision making", *ASCE, J. Constr. Engrg. and Mgmt.*, 127 (1), 35-45.

Chua, D.K.H. and Goh, Y.M. (2004a). "Utilising the modified loss causation model for the codification and analysis of accident data". In *Construction Safety Management Systems*, Rowlinson, S. (ed.), pp. 443-464. Spon Press, London.

Chua, D.K.H. and Goh, Y.M. (2004b). "An incident causation model for improving feedback of safety knowledge", *ASCE, J. Constr. Engrg. and Mgmt*., 130 (4), 542-551.

Chua, D.K.H. and Goh, Y.M. (2004c). "A Poisson model of construction incident occurrence", *ASCE, J. Constr. Engrg. and Mgmt*. (Accepted for publication).

Cognitive Systems Inc. (1992). *Remind: Solutions from prior experience- Developer's Reference Manual*. Cognitive Systems Inc., Boston.

Conover, W. J. (1980). *Practical non-parametric statistics*, 2nd Edition. John Wiley, New York.

Covey, S. R. (1989). *The 7 Habits of Highly Effective People*. Simon & Schuster UK Ltd, London.

Cox, D. R. and Lewis, P. A. (1966). *The Statistical analysis of Series of events*. Methuen and Wiley, London.

CPWR (1993). "An agenda for change", *Report of the National Conference on Ergonomics, Safety, and Health in Construction.* Center to Protect Workers' Rights, Washington DC.

Crawley, F., and Tyler, B. (2003). *Hazard identification method.* Institution of Chemical Engineers, Rugby.

Cunningham, P. and Bonzano, A. (1999). "Knowledge engineering issues in developing a case-based reasoning application", *Knowledge-Based Systems*, 12, 371-379.

Davenport, T.H. (1997). "Ten principles of knowledge management and four case studies", *Knowledge and Process Management*, 4(3), 187-208.

Davenport, T.H. and Beers, M.C. (1995). "Managing information about processes", *Journal of Management Information Systems*, 12(1), 57-80.

Domeshek, E. (1991b). "Indexing stories as social advice." In *Proc. AAAI-91*. AAAI Press/MIT Press, Cambridge, MA.

Dutton, D. and Maun, K. (1997). "Case-based reasoning for materials selection in design", *International Journal of Construction Information Technology,* 5(2), 103–111.

Dym, C. L. and Levitt, R. E. (1991). *Knowledge Based Systems in Engineering.* McGraw-Hill, New York.

Empolis Knowledge Management GmbH (2001). *cbr:works 4- Compendium* Available at: http://www.empolis.com.

Everett, J.G. and Frank, P.B.J. (1996). "Costs of accidents and injuries to the construction industry", *ASCE, J. Constr. Engrg. and Mgmt*., 122(2), 158-.

Feyer, A. and Williamson, A. M. (1991). "A classification system for causes of occupational accidents for use in preventive strategies", *Scandinavian Journal of Work, Environment and Health*, 17, 302-311.

Fowler, A. (2000). "The role of AI-based technology in support of the knowledge management value activity cycle", *Journal of Strategic Information Systems*, 9, 107-128.

Fridstrom, L., Ifver, I., Ingebrigtsen, S., Kulmala, R., and Thomsen, L. K. (1995). "Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accident counts", *Accident analysis and prevention*, 27 (1), pp. 1-20.

Gero, J.S. (1990). "Design prototypes: a knowledge representation schema for design", *AI Magazine*, 11(4), 26–36.

Goh, Y.M. and Chua, D.K.H. (2002). "Identification of Factors Causing Fatal Construction Accidents". In *Proc. CIB W99 3rd international conference on implementation of safety and health on construction sites,* Rowlinson, S. (Ed), pp. 69-76. University of Hong Kong, Hong Kong.

Gordon, R.P.E. (1998). "The contribution of human factors to accident in the offshore oil industry", *Reliability Engineering and System Safety*, 61, 95-108.

Haddon, W. (1980). "The basic strategies for reducing damage from hazards of all kinds." *Hazard Prevention*, September/October: 8-12.

Hadikusumo B. H. W. and Rowlinson, S. (2002). "Integration of virtually real construction model and design-for-safety-process database", *Automation in Construction*, 11(5), 501-509.

Hammer, M. and Champy, J. (1993). *Reengineering the Corporation*. N. Brealey, London.

Hammond, K. (1986a). “CHEF: A model of case-based planning.” In *Proc. of AAAI-86.* AAAI Press/MIT Press, Cambridge.

Han, J. W. and Kamber, M. (2001). *Data Mining- Concepts and Techniques.* Morgan Kaufmann Publishers, London.

Harms-Ringdhal, L. (1993). *Safety Analysis: Principles and practice in occupational safety.* Elsevier Applied Science, London.

Hegazt T., Moselhi, O. and Fazio, P. (1994). “Developing practical neural network applications using backpropagation”, *J. Microcomputers in Civ. Engrg*., 9(2), 145-159.

Hegazy, T. and Ayed, A. (1998). “Neural network model for parametric cost estimation of highway projects”, *ASCE, J. Constr. Engrg. and Mgmt*., 124(3), 210-218.

Heinrich, H. W. (1939). *Industrial Accident Prevention*. McGraw-Hill, New York.

Henderson, J., Whittington, C. and Wright, K. (2001). *Accident investigation – The drivers, methods and outcomes*, HSE CRR 344/2001. HSE, London.

Hinrichs, T. (1992). *Problem solving in open worlds: A case study in design.* Erlbaum, Northvale, NJ.

Hinze, J. (1997). *Construction safety*. Prentice Hall, New Jersey.

Hinze, J., Pedersen, C. and Fredley, J. (1998). “Identifying root causes of construction injuries”, *ASCE, J. Constr. Engrg. and Mgmt*., 124(1), 67-71.

Holland J.H. (1986). “Escaping brittleness: the possibility of general purpose learning algorithm applied to parallel rule-based system”, In *machine learning- A guide to current research*, Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds), pp. 199-223. Kulwer Academic Publichers, Chicago.

Illingworth, J. R. (2001). *Construction Methods and Planning*. Spon Press, London.

Jeong, B.Y. (1998). “Occupational deaths and injuries in the construction industry”, *Applied Ergonomics*, 29(5), 355-360.

Johnson, W. G. (1980). *MORT Safety Assurance System*. Marcel Dekker, New York.

Kang, B., Lee, B., Kang, K. Y., Suh, J. C. and Yoon, E. S. (1999). “AHA: a knowledge based system for automatic hazard identification in chemical plant by multimodel approach”, *Expert Systems with Application*, 16, 183-195.

Karim, A. and Adeli, Hojat (2003). “CBR model for freeway work zone traffic management”, *Journal of Transportation Engineering*, 129(2), 134-145.

Kartam, N. A. (1997). “Integrating safety and health performance into construction CPM”, *ASCE, J. Constr. Engrg. and Mgmt*., 121(2), 121-126.

Kartam, N.A. and Bouz, R.G. (1998). “Fatalities and injuries in the Kuwaiti construction industry”, *Accident Analysis and Prevention*, 30 (6), 805-814.

Kasabov, N.K. (1999). *Foundations of neural networks, fuzzy systems, and knowledge engineering*. NetLibrary Inc., Boulder.

Kerr, W. (1950). “Accident proneness of factory departments.”, *J. Applied Psychology*, 34, 167-170.

Kerr, W. (1957). “Complementary theories of safety psychology”, *J. Social Psychology*, 43, 3-9.

Kjellén (2000). *Prevention of accidents through experience feedback.* Taylor & Francis, New York.

Kjellèn U. (1987). “Simulating the use of a computerized injury and near accident information system in decision making”, *Journal of Occupational Accidents,* 9, 87-105.

Kletz, T. (1994). *Learning from accidents,* 2nd ed., Butterworth-Heinemann, Boston.

Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann Publisher, San Mateo, CA.

Kolodner, J. L. (1984). *Retrieval and organisation strategies in conceptual memory: A computer model.* Erlbaum, NJ.

Kolodner, J.L. (1988). “Retrieving events from a case memory: A parallel implementation”. In *Proc. Workshop on case-based reasoning (DARPA),* Clearwater, Florida, Kolodner, J. L. (ed). Morgan Kaufmann, San Mateo, CA.

Kolodner, J.L. and Simpson, R.L. (1989). “The MEDIATOR: Analysis of an early case-based problem solver”, *Cognitive Science,* 13(4), 507-549.

Koton, P. (1989). *Using experience in learning and problem solving*, Ph.D. diss., Department of Computer Science, MIT.

Kumamoto, H. and Henley, E. J. (1996). *Probabilistic Risk Assessment and Management for Engineers and Scientist,* 2nd Edition. IEEE Press, New York.

Larsson, T. J. and Field, B. (2002). “The distribution of occupational injury risks in the Victorian construction industry”, *Safety Science*, 40, 439-456.

Li, H. (1996). “Selecting KBES development techniques for applications in the construction industry”, *Construction Management and Economics,* 14, 67-74.

Liao, T. W., Zhang, Z. M. and Mount, C. R. (1998). “Similarity measures for retrieval in Case-Based Reasoning Systems”, *Applied Artificial Intelligennce*, 12, 267-288.

Liao, T. W., Zhang, Z. M. and Mount, C. R. (2000). “A case-based reasoning system for identifying failure mechanisms”, *Engineering Application of Artificial Intelligence*, 13, 199-213.

Livingston, A. D., Jackson, G. and Priestley, K. (2001). *Root causes analysis: Literature review*. HSE, London.

Luu, D. T., Ng, S. T., Chena, S. E. (2003). “A case-based procurement advisory system for construction”, *Advances in Engineering Software*, 34, 429–438.

Maher, M.L. and Garza, A.G. (1996). “Developing case-based reasoning for structural design”, *IEEE Expert*, 42–52.

McKinnon, R. C. (2000). *Cause, effect, and control of accidental loss with accident investigation kit*. Lewis Publishers, Boca Raton.

McQueen, R.J. and Kock, N.F. (1996). “Learning and process improvement in the knowledge engineering organisation: a critical analysis of four contemporary myths”, *The Learning Organisation*, 3(1), 31-41.

Modarres, M., Kaminskiy, M. and Krivtsov, V. (1999). *Reliability engineering and risk analysis- A practical guide.* Marcel Dekker Inc, New York.

Mohan, S. (1990). “Expert systems application in construction management and engineering”, *ASCE, J. Constr. Engrg. and Mgmt*., 116(1), 87-99.

MOM (2001). “Occupational safety statistics”, *Ministry of Manpower official website*. Available at: <http://www.gov.sg/mom/menubar/public/pmain.htm>. Accessed September 2001.

Montgomery, D. C. and Runger, G. C. (1999). *Applied Statistics and Probability for Engineers*, 2nd Edition. John Wiley, Singapore.

Mount, C. and Liao T. W. (2001). “Prototype of an intelligent failure analysis system”, In *Proc. ICCBR 2001, LNAI 2080*, Aha D. W. and Watson I. (Eds.), 716-730. Springer-Verlag, Berlin Heidelberg.

MSHA (2004). US Department of Labor, Mine Safety and Health Administration. Available at: http://www.msha.gov/TRAINING/PART46/PT46TRAIN.HTM. Accessed May 2004.

National Safety Council (1997). Accident facts. Itasca, Ill.

Ng, S.T., Smith, N.K. and Skitmore, R.M. (1998). “A case-based reasoning model for contractor prequalification”, *International Journal of Construction Information Technology* 6(1), 47–61.

Nicholson, A. J. (1985). “The variability of accident counts”, *Accident analysis and prevention*, 17, 47-56.

Nicholson, A. J. (1986). “The randomness of accident counts”, *Accident analysis and prevention*, 18, 193-198.

Nicholson, A. J. and Wong, Y. D. (1993). “Are accidents Poisson distributed? A statistical test”, *Accident analysis and prevention*, 25(1), 91-97.

Ong, L. S. (2002). *Case-Based Reasoning*, Course notes, Institute of Systems Science, National University of Singapore, Singapore.

Park, C.S. and Han, I. (2002). “A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction”, *Expert Systems with Applications*, 23(3), 255-264.

Pearce, M., Goel, A.K., Kolodner, J., Zimring, C., Sentosa, L. and Billington, R.(1992). “Case-based decision support: a case study in architectural design”, *IEEE Expert* 7, 5, 14–20.

PSB (Singapore Productivity and Standards Board) (1999). *CP79:1999- Code of Practice for Safety management system for construction worksites*. PSB, Singapore.

Ramakrishnan, R. and Gerhrke, J. (2000). *Database Management Systems*, 2nd Edition. McGraw-Hill Higher Education, Singapore.

Ramsey, J. (1985). “Ergonomics factors in task analysis for consumer product safety”. *Journal of Occupational Accidents*, 7, 113-123.

Rasmussen, J. (1982). “A taxonomy for describing human malfunction in industrial installations”, *Journal Of Occupational Accidents*, 4, 311-333.

Reason, J. (1990). *Human Error*. Cambridge University Press, London.

Reason, J. (1993). “Managing the management risk: new approaches to organisational safety.” *Reliability and safety in hazardous work systems*, B. Wilpert and T. Qvale, (eds.), pp. 7-22. Lawrence Erlbaum Associates Ltd, London.

Redmill, F., Chudleigh, M., and Catmur, J. (1999). *System safety: HAZOP and Software HAZOP*, John Wiley, England.

Redmond, M.A. (1990a). “Distributed cases for case-based reasoning: Facilitating use of multiple cases.” In *Proc. AAAI-90.* AAAI Press/MIT Press, Cambridge, MA.

Redmond, M.A. (1992). “Learning by observing and understanding expert problem solving”. In *Proc. Thirteenth Annual Conference of the Cognitive Science Society*. Erlbaum, Northvale, NJ.

Ross, S. M. (2000). *Introduction to probability models*. Academic Press, Florida.

Russel, S.J. and Norvig, P. (1995). *Artificial intelligence : a modern approach*. Prentice Hall, Englewood Cliffs, N.J.

Russell, S and Norvig, P (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence. Englewood Cliffs, New Jersey.

Sanders, M., and Shaw, B. (1988). *Research to determine the contribution of system factors in the occurrence of underground injury accidents*. Bureau of Mines, Pittsburgh, PA.

Sawacha, E., Naoum, S. and Fong, D. (1999). "Factors affecting safety performance on construction sites", *International Journal of Project Management*, 17(5), 309-315.

Simpson, R. L. (1985). *A computer model of case-based reasoning in problem solving: An investigation in the domain of dispute mediation.* Georgia Institute of Technology, School of Information and Computer Science Technical Report no. GIT-ICS-85/18. Atlanta.

Sinha, A. P., and May, J. H. (2001). "Providing design assistance: a case-based approach", *Information Systems Research*, 7(3), 363-387.

Suh, J. C., Lee, S., and Yoon, E. S. (1997a). "New strategy for automated hazard analysis of chemical plants. Part 1: knowledge modeling", *Journal of Loss Prevention in the Process Industries*, 10 (2), 113-126.

Suh, J. C., Lee, S., and Yoon, E. S. (1997b). "New strategy for automated hazard analysis of chemical plants. Part 2: reason algorithm and case study", *Journal of Loss Prevention in the Process Industries*, 10(2), 127.

Suh, M. S., Jhee, W. C., Ko, Y. K., Lee, A. (1998). "A case-based expert system approach for quality design", *Expert Systems with Application*, 15, 181-190.

Suraji, A., Duff, A. R., and Peckitt, S. J. (2001). "Development of causal model of construction accident causation" *ASCE, J. Constr. Engrg. and Mgmt.*, 127(4), 337-344.

Swain, A. and Guttmann, H. (1983). *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications (NUREG/CR-1278)*, Nuclear Regulatory Commission, Washington DC.

Sycara, E.P. (1987). *Resolving adaversarial conflicts: An approach to integrating case-based and analytic methods.* Georgia Institute of Technology, School of Information and Computer Science Technical Report no. GIT-ICS-87/26. Atlanta.

Tah, J.H.M. and Howes, R. (1998). "An application of case-based reasoning to the planning of highway bridge construction", *Engrg. Construct. Architectural Mgmt.*, 5(4), 327-338.

Tam, C.M. and Fung, I. W. H. (1998). "Effectiveness of safety management strategies on safety performance in Hong Kong." *Const. Mgmt. and Economics.* 16(1), 49-55.

Tan, W. (2004). *Practical research methods*. Prentice Hall, Singapore.

Thomson ISI (2003). *Web of Science- Science Citation Index (expanded).* Available at:

www.isinet.com.

Vaidhyanathan, R. and Venkatasubramanian, V. (1996). "Experience with an expert system for automated HAZOP analysis", *Computers and Chemical Engineering*, 20 (suppl.), S1589-S1594.

Vong, C.M., Leung, T.P. and Wong, P.K. (2002). "Case-based reasoning and adaptation in hydraulic production machine design", *Engineering Applications of Artificial Intelligence*, 15, 567–585.

Watson, I. (1997). *Applying Case-based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufmann Publishers, San Francisco.

Weatherill, T. and Cameron, I. T. (1989). "A prototype expert system for hazard and operability studies", *Computers and Chemical Engineering*, 13 (11/12), 1229-1234.

Whittington, C., Livingston, A., and Lucas, D. A. (1992). *Research into management, organizational and human factors in the construction industry*, HSE CRR No. 45/1992. HSE Books, Sudbury, Suffolk.

Wig, K.M. (1993). *Knowledge Management Foundations, Thinking About Thinking*, Lexington Books, Massachusetts.

Wolff, R. W. (1989). Stochastic modeling and the theory of queues. Prentice Hall, Englewood Cliffs, N.J.

Yang, S., Robertson, D. (1994). "A case-based reasoning system for regulatory information", In *Proc. IEE Colloq. on Case-Based Reasoning: Prospects for Applications*, Digest No. 1994/057, pp. 3/1–3/3.

Yau, N. J. and Yang J. Y. (1998). "Applying case-based reasoning technique to retaining wall selection", *Automation in Construction*, 7(4), 271-283.

Zhang, Z.-M. (1998). *Applying case-based reasoning and genetic algorithm to failure analysis*, MS Thesis, Louisana State University.

LIST OF PUBLICATIONS

Book Chapter

Chua, D.K.H. and Goh, Y.M. (2004a). "Utilising the Modified Loss Causation Model for the Codification and Analysis of Accident Data". In *Construction Safety Management Systems*, Rowlinson, S. (ed.), pp. 443-464. Spon Press, London.

Journal Papers

Chua, D.K.H. and Goh, Y.M. (2004b). "An incident causation model for improving feedback of safety knowledge", *ASCE, J. Constr. Engrg. and Mgmt.*, 130 (4), 542-551.

Chua, D.K.H. and Goh, Y.M. (2004c). "A Poisson Model of Construction Incident Occurrence", *ASCE, J. Constr. Engrg. and Mgmt.* (Accepted for publication).

Goh, Y.M. and Chua, D.K.H. (2003a). "A construction safety knowledge database", *Journal of the Institution of Engineers, Singapore*, 43 (3), 29-37.

Goh, Y.M. and Chua, D.K.H. (2004a). "A CBR Approach to Construction Safety Risk Assessment: Knowledge Representation and Case Retrieval" (to be submitted; content of paper based on chapters 4 and 6).

Goh, Y.M. and Chua, D.K.H. (2004b). "A CBR Approach to Construction Safety Risk Assessment: Adaptation and Utilisation of Cases" (to be submitted; content of paper based on chapters 5 and 6).

Conference Papers

Chua, D.K.H. and Goh, Y.M. (2002). "Application of Case Based Reasoning in Construction Safety Planning". In *Proc. Computing in Civil Engineering, International Workshop on IT in Civil Engineering*, Songer, AD and Miles, JC (Eds), pp. 298-307. ASCE, Washington D.C.

Goh, Y.M. and Chua, D.K.H. (2002). "Identification of Factors Causing Fatal Construction Accidents". In *Proc. CIB W99 3rd international conference on implementation of safety and health on construction sites*, Rowlinson, S. (Ed), pp. 69-76. University of Hong Kong, Hong Kong.

Goh, Y.M. and Chua, D.K.H. (2003b). "Improving Safety Risk Analysis Through The Application Of Case Based Reasoning Concepts". In *Proc. Innovation in Architecture, Engineering and Construction (AEC) 2003*, CJ Anumba (Ed), pp. 441-451, Loughborough University, Loughborough.

Appendix 1

THE MODIFIED LOSS CAUSATION MODEL (MLCM)

TAXONOMY

Table A1.1 Taxonomy for Situational Variables (Type of work)

1. Situational Variables- Type of Work	
1.1 Architectural/Renovation/Finishing work 1.1.1 Roofing work 1.1.2 Finishing work 1.1.3 Plastering 1.1.4 Painting 1.1.5 Installation of non-structural component 1.1.6 Other A/R/F work	1.2 Building services work 1.2.1 Electrical work 1.2.2 Piping work 1.2.3 Air-con dismantling/installation 1.2.4 Other building services work
1.3 Geotechnical work 1.3.1 Excavation work 1.3.2 Trench work 1.3.3 Tunnelling work 1.3.4 Piling work 1.3.5 Other Geotechnical work	1.4 Material/equipment handling/transportation 1.4.1 Operation of vehicle/transport 1.4.2 Lifting/lowering 1.4.3 By Crane 1.4.4 By other equipment/plant 1.4.5 Manual handling 1.4.6 Other Material/equipment H/T
1.5 Plant/ machinery/ equipment maintenance/ dismantling /installation 1.5.1 Dismantling of P/M/E 1.5.2 Servicing of P/M/E 1.5.3 Installation of P/M/E	1.6 Structural work 1.6.1 Concreting 1.6.2 Installation of precast components 1.6.3 Demolition of structural components 1.6.4 Erection/dismantling of formwork/falsework 1.6.5 Erection/dismantling of lifting platform 1.6.6 Erection/dismantling of mobile working platform 1.6.7 Erection/dismantling of temporary access scaffold 1.6.8 Other structural work
1.7 Other types of work 1.7.1 Marine construction 1.7.2 Reclamation work 1.7.3 Housekeeping work 1.7.4 Movement around site 1.7.5 Other operations of vehicle not elsewhere classified	

Table A1.2 Taxonomy for Types of Contact Event

2. Types of Contact Event	
2.1 Fall of person 2.1.1 Struck ground 2.1.2 Struck sharp object 2.1.3 Struck other objects	2.2 Struck by falling objects 2.2.1 Earth, rocks, stones etc. 2.2.2 Structure (parts of building, temporary structure, etc) 2.2.3 Equipment 2.2.4 Construction materials 2.2.5 Other struck by falling objects
2.3 Striking against or struck by objects 2.3.1 Striking against stationary objects (excluding fall of persons) 2.3.2 Striking against moving objects (excluding fall of persons) 2.3.3 Struck by moving objects 2.3.3.1 Flying fragments and other small objects 2.3.3.2 Plant/ machinery/ equipment/ vehicle 2.3.3.3 Lifted object 2.3.3.4 Other moving objects 2.3.4 Cut by object 2.3.5 Other striking against or struck by objects	2.4 Caught in or between objects 2.4.1 Caught in an object 2.4.1.1 Plants/vehicles 2.4.1.2 Other objects 2.4.2 Caught between stationary object and moving object 2.4.3 Caught between moving objects 2.4.4 Other caught in or between objects
2.5 Over-exertion or strenuous movements 2.5.1 Over-exertion in lifting objects 2.5.2 Over-exertion in pulling or pushing objects 2.5.3 Over-exertion in handling or throwing objects 2.5.4 Other over-exertion or strenuous movements	2.6 Exposure/contact with extreme temp/pressure 2.6.1 Exposure to high heat atmosphere or environment (excluding fire/explosion) 2.6.2 Exposure to cold atmosphere or environment 2.6.3 Contact with extreme hot substances or objects 2.6.4 Contact with extreme cold substances or objects 2.6.5 Other exposure/contact with extreme temperature/pressure
2.7 Exposure/contact with electric current	2.8 Exposed to harmful substances/radiations 2.8.1 Contact by inhalation, ingestion or absorption of harmful substances 2.8.2 Exposure to radiations 2.8.3 Contact with corrosive substances 2.8.4 Other type of exposure to harmful substances/radiations
2.9 Other types of incidents 2.9.1 Drowning 2.9.2 Other types of incidents not elsewhere classified 2.9.3 Unclassifiable incidents due to lack of information	

Table A1.3 Taxonomy for Types of Breakdown Event

3. Types of Breakdown Event	
3.1 Collapse/toppling of object 3.1.1 Plant/machinery (including parts of machinery)/equipment 3.1.2 Soil structure (Earth, rocks, stones etc.) 3.1.3 Structure under work 3.1.4 Temporary structures 3.1.4.1 Access scaffold 3.1.4.1.1 Non-mobile access scaffold 3.1.4.1.2 Mobile access scaffold 3.1.4.1.3 Working platforms on access scaffold 3.1.4.1.4 Other access scaffold 3.1.4.2 Formwork/falsework 3.1.4.3 Lifting platform 3.1.4.4 Working platforms (excluding working platforms on access scaffold) 3.1.4.5 Other temporary structure 3.1.5 Other types of collapses	3.2 Lost of balance- Fall of person 3.2.1 Slipped 3.2.2 Stepped into space 3.2.3 Stepped on fragile material 3.2.4 Tripped 3.2.5 Other types of lost of balance 3.2.6 Unknown type of lost of balance 3.3 Object fall off surface 3.3.1 Object slipped off surface 3.3.2 Object under hoist dislodged 3.3.3 Object fall into depth 3.3.3.1 Soil structure/ excavation 3.3.3.2 Manhole 3.3.3.3 Other types of depth 3.3.4 Other types of fall off surface 3.4 Loss control of plant/vehicle (Runaway plant/vehicle) 3.5 Collision between objects 3.5.1 Lifted objects 3.5.2 Plant/vehicles 3.5.3 Other moving objects
3.6 Failure of equipment (breakage)	3.7 Fire/explosion
3.8 Other types of breakdown event	

Table A1.4 Taxonomy for Types of Substandard Physical Conditions (Immediate Causes)

4. Types of Substandard Physical Conditions (Immediate Causes)	
4.1 Substandard plant/machinery/equipment/tools 4.1.1 Defective plant/machinery/equipment/tools 4.1.2 Lack of proper safety feature 4.1.3 Other substandard plant/machinery/equipment/tools	4.2 Substandard construction material 4.2.1 Improper chemical composition 4.2.2 Other substandard construction material
4.3 Substandard structures/parts of structure 4.3.1 Lack of proper safety structure 4.3.2 Insufficient structural capacity 4.3.3 Defective/damaged structure/parts of structure 4.3.4 Other substandard structures/parts of structure	4.4 Substandard work environment 4.4.1 Slippery conditions 4.4.2 Tripping conditions 4.4.3 Congestion/restrictive conditions 4.4.4 Poor weather conditions 4.4.5 Lack of insulation against high energy source 4.4.5.1 Electrical energy 4.4.5.2 Heat energy 4.4.5.3 Lack of heat energy 4.4.5.4 Other sources of energy 4.4.6 Poor ventilation 4.4.7 Lack of proper warning signs/signals 4.4.8 Other substandard work environment conditions
4.5 Other substandard physical condition	

Table A1.5 Taxonomy for Types of Substandard Acts (Immediate Causes)

5. Types of Substandard Acts (Immediate Causes)	
5.1 Extraneous Acts 5.1.1 Horseplay 5.1.2 Under influence of alcohol/drugs 5.1.3 Other extraneous acts	5.2 Improper equipment usage 5.2.1 Inappropriate activation of control 5.2.2 Making safety device inoperative 5.2.3 Servicing equipment in operation 5.2.4 Using defective equipment 5.2.5 Using equipment for inappropriate purpose/situation 5.2.6 Lack of control of equipment or machinery 5.2.7 Using the right equipment but in the wrong manner 5.2.8 Other improper equipment usage
5.3 Inappropriate response to emergency 5.3.1 Inappropriate emergency response 5.3.2 Inappropriate response to prevent incident 5.3.3 Other inappropriate response to emergency	5.4 Omission of basic safety measures 5.4.1 Failure to check 5.4.2 Failure to secure/back-up 5.4.3 Failure to use PPE 5.4.3.1 Failure to use 5.4.3.2 Failure to use in proper manner 5.4.4 Failure to warn 5.4.5 Other omission of basic safety measures
5.5 Spatial error 5.5.1 Failure to use proper access/egress 5.5.2 Improper placement of objects 5.5.3 Improper position/location for task 5.5.4 Other spatial errors	5.6 Improper work procedure 5.6.1 Failure to perform a procedure/steps of a procedure 5.6.2 Mismatch of workers' capacity and demands of procedure 5.6.3 Operating without proper authority/permission 5.6.4 Perform procedure in the wrong sequence 5.6.5 Perform wrong/inappropriate procedures/steps 5.6.6 Other improper work procedures
5.7 Other substandard acts	

Table A1.6 Taxonomy for Types of Personal Factors (Immediate Causes and Underlying Factors)

6. Types of Personal Factors (Immediate Causes and Underlying Factors)	
6.1 Lack of knowledge/skill 6.1.1 Lack of experience/practice/performance 6.1.2 Inadequate orientation 6.1.3 Inadequate initial/update training 6.1.4 Other lack of knowledge/skill	6.2 Mental/psychological factors 6.2.1 Emotional factors 6.2.2 Mental fatigue 6.2.3 Inadequate mental capability 6.2.4 Poor judgement 6.2.5 Confusing instructions 6.2.6 Distracting events 6.2.7 Others
6.3 Improper motivation 6.3.1 Mismatch of safe performance and reward/punishment 6.3.2 Excessive frustration 6.3.3 Inappropriate attempt to save time of effort or avoid discomfort 6.3.4 Inadequate discipline 6.3.5 Inappropriate peer pressure 6.3.6 Improper supervisory example 6.3.7 Inadequate performance feedback 6.3.8 Others	6.4 Physical/physiological factors 6.4.1 Inadequate physical/physiological capabilities 6.4.2 Physical fatigues 6.4.3 Injury/illness 6.4.4 Other physical/physiological factors
6.5 Other personal factors	

Table A1.7 Taxonomy for Types of SMS Failures (PSB 1999)

7. Types of SMS Failures (A) Lack of measure (B) Inadequate measure (C) Inadequate execution	
7.1 Safety policy 7.1.1 Safety Organisation 7.1.2 Policy Review	7.2 Safe work practices 7.2.1 Application of Safe work practices 7.2.2 Permit-to-work system 7.2.3 Statutory requirements
7.3 Safety training 7.3.1 Identification of training needs 7.3.2 Training for management personnel 7.3.3 Training for supervisory personnel 7.3.4 Training for workers 7.3.5 Training records	7.4 Group meetings 7.4.1 Safety committee meeting 7.4.2 Tool box meetings and safety briefings 7.4.3 Coordination meeting
7.5 Incident investigation and analysis 7.5.1 Identification and record of incidents 7.5.2 Investigation of incidents 7.5.3 Analysis of incident statistics	7.6 In-house safety rules and regulations 7.6.1 In-house rules and regulations 7.6.2 Training and review of rules and regulations 7.6.3 Safety sign
7.7 Safety promotion 7.7.1 Promotional activities 7.7.2 Safety bulletin boards 7.7.3 Recognition of good safety performance 7.7.4 Records of promotion activities	7.8 Evaluation, selection and control of sub-contractors 7.8.1 Evaluation of sub-contractors 7.8.2 Selection of sub-contractors 7.8.3 Control of sub-contractors
7.9 Safety inspections 7.9.1 Competency of safety inspectors 7.9.2 Inspection methodology 7.9.3 Follow-up system	7.10 Maintenance regime for all machinery and equipment 7.10.1 Maintenance program 7.10.2 Competency of maintenance program
7.11 Hazard analysis 7.11.1 Hazard analysis plan 7.11.2 Hazard analysis method 7.11.3 Hazard analysis report	7.12 The control of movements & use of haz. subst. & chem. 7.12.1 Management of hazardous substances and chemicals
7.13 Emergency preparedness 7.13.1 Emergency plan 7.13.2 Emergency team 7.13.3 Emergency drills and exercises	7.14 Occupational health program 7.14.1 Hearing conservation program 7.14.2 Respiratory protection program 7.14.3 Training and education

Table A1.8 Taxonomy for Types of Job Factors (Underlying Factors)

8. Types of Job Factors (Underlying Factors)	
8.1 Factors related to designers 8.1.1 Inadequate structural capacity 8.1.2 Lack of consideration for site safety 8.1.3 Lack of communication of design to Site Management 8.1.4 Other factors related to designers	8.2 Factors related to operatives 8.2.1 Non-compliance to SM's instructions 8.2.2 Poor communication between operatives 8.2.3 Failed to perform to standard or expected competency 8.2.4 Failure to feedback to SM on problems 8.2.5 Other factors related to operatives
8.3 Factors related to project management/corporate 8.3.1 Lack of commitment to safety 8.3.2 Lack of financial support for safety efforts 8.3.3 Lack of communication of safety priority to site management 8.3.4 Lack of audit on site safety 8.3.5 Other factors related to project management/corporate	8.4 Factors related to site management 8.4.1 Failure to construct according to designers'/manufacturers' design 8.4.2 Failure to identify /analyse hazards 8.4.3 Failure to manage identified unacceptable hazards 8.4.4 Failure to obtain/allocate adequate/proper resources 8.4.4.1 Human resources 8.4.4.2 Physical resources 8.4.4.3 Other resources 8.4.5 Failure to set up proper safe work procedures 8.4.6 Failure to communicate safe work procedures/hazards 8.4.7 Inadequate supervision of site activities 8.4.7.1 Failure to monitor site progress 8.4.7.2 Failure to ensure compliance to safe work procedures 8.4.7.3 Inadequate inspection of constructed components 8.4.8 Violation of safe work procedures 8.4.9 Failure to delegate/coordinate site work 8.4.10 Lack of proper maintenance of physical resources 8.4.10.1 Temporary structures 8.4.10.2 Plant 8.4.10.3 Equipment 8.4.10.4 Other physical resources 8.4.11 Other factors related to site management 8.4.11.1 Over-emphasis on production goals 8.4.11.2 Others
8.5 Other job factors 8.5.1 Poor work communication between jobs 8.5.2 Other job factors not elsewhere classified	

Table A1.9 Taxonomy for Types of Organisational Factors (Underlying Factors)

9. Types of Organisational Factors (Underlying Factors)	
9.1 Poor safety and/ or organisational culture	9.2 Inappropriate organisational structure
9.3 Lack of organisational learning	9.4 Lack of stable workforce
9.5 Lack of formal and informal communication structure	9.6 Other organisational factors

Appendix 2

STATISTICAL RESULTS OF ANALYSIS ON 140 FATAL ACCIDENTS

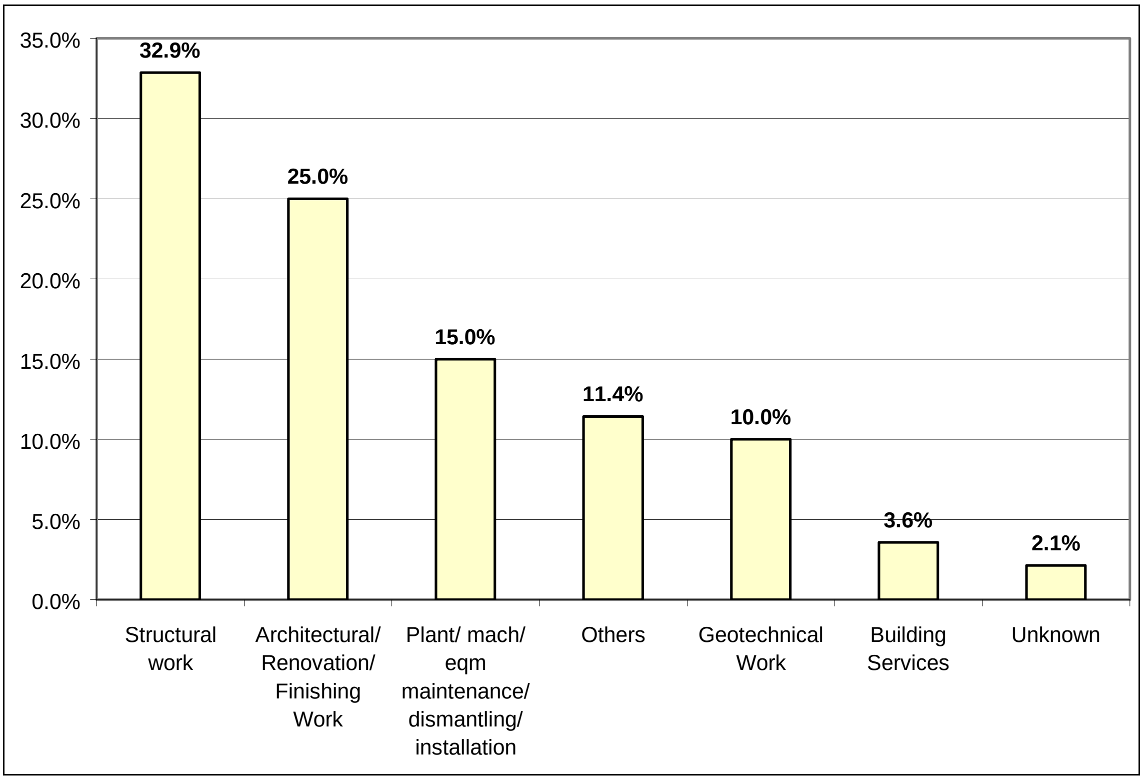


Figure A2.1 Distribution of type of work

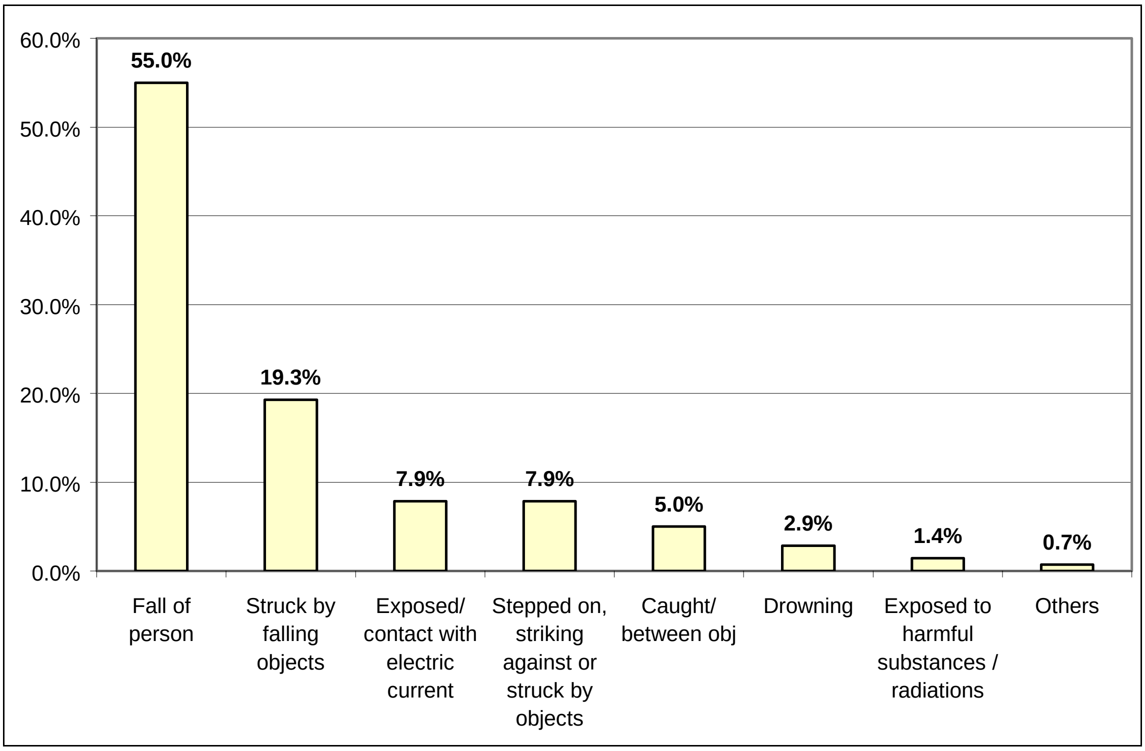


Figure A2.2 Distribution of type of contact event

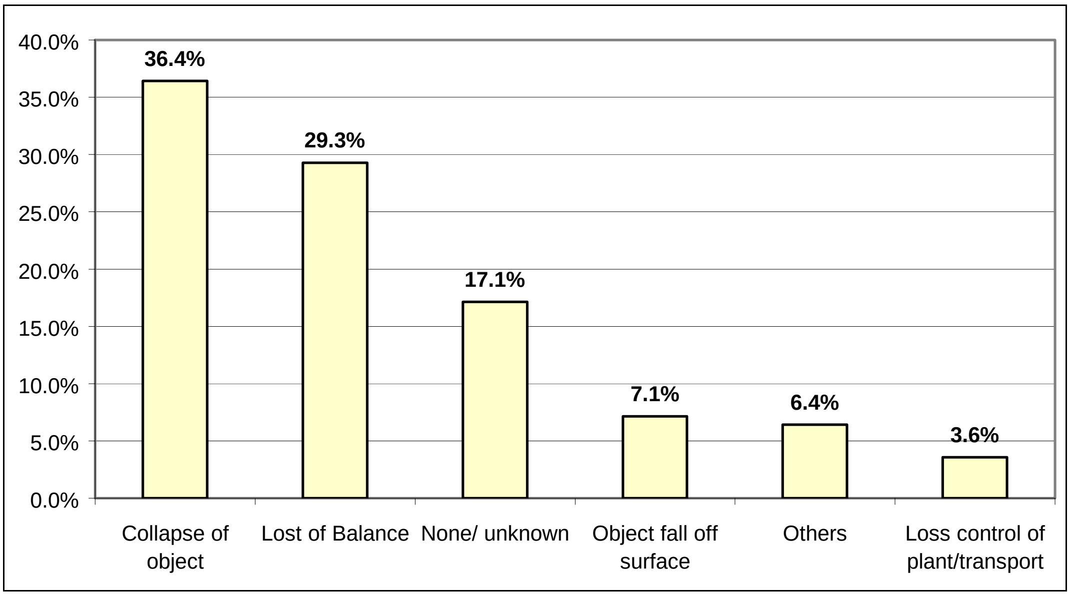


Figure A2.3 Distribution of type of breakdown event

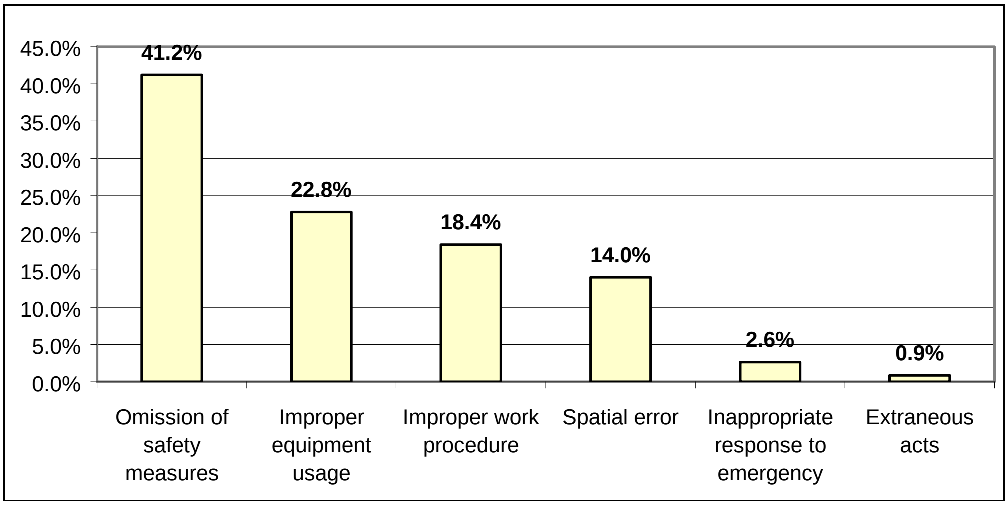


Figure A2.4 Distribution of type of substandard acts

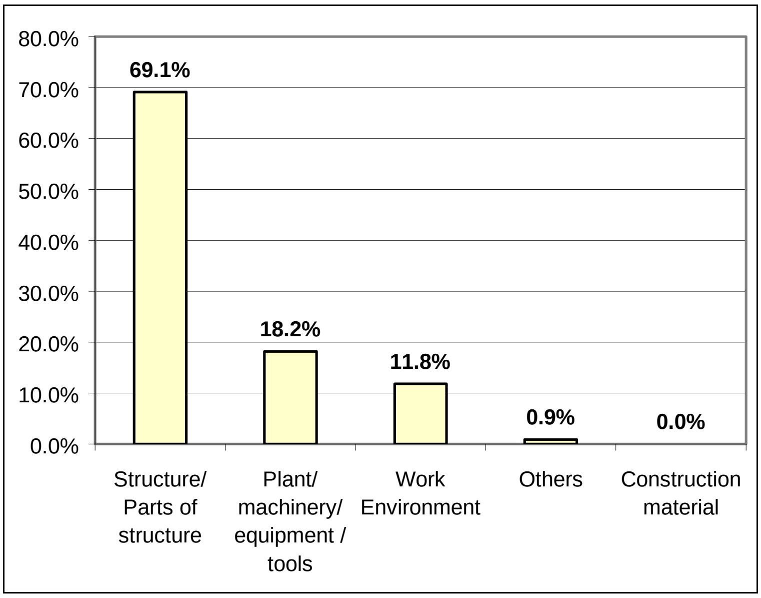


Figure A2.5 Distribution of type of substandard physical conditions

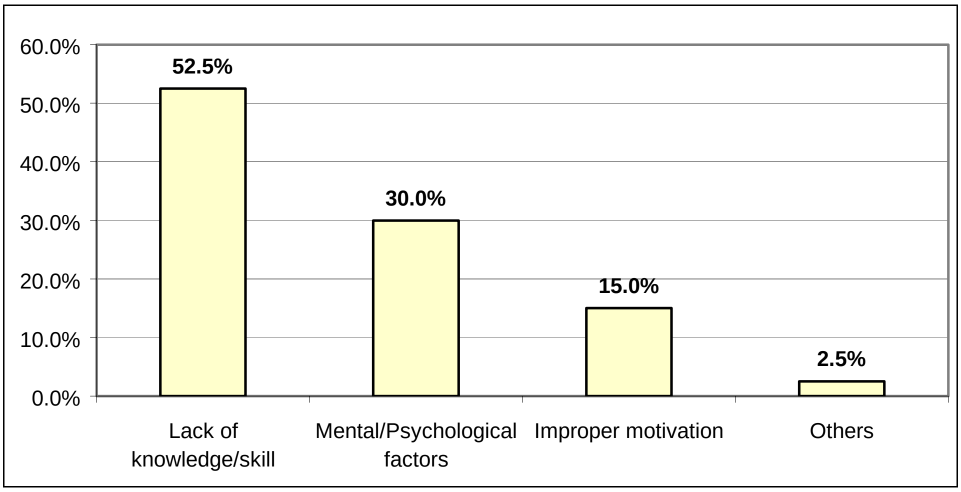


Figure A2.6 Distribution of types of immediate personal factors

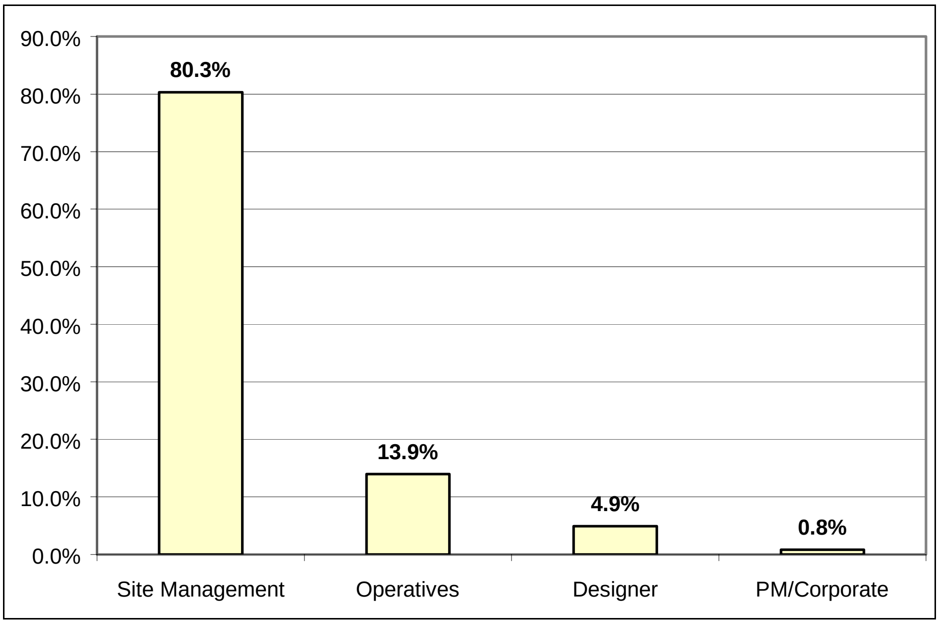


Figure A2.7 Distribution of type of job factors base on job function

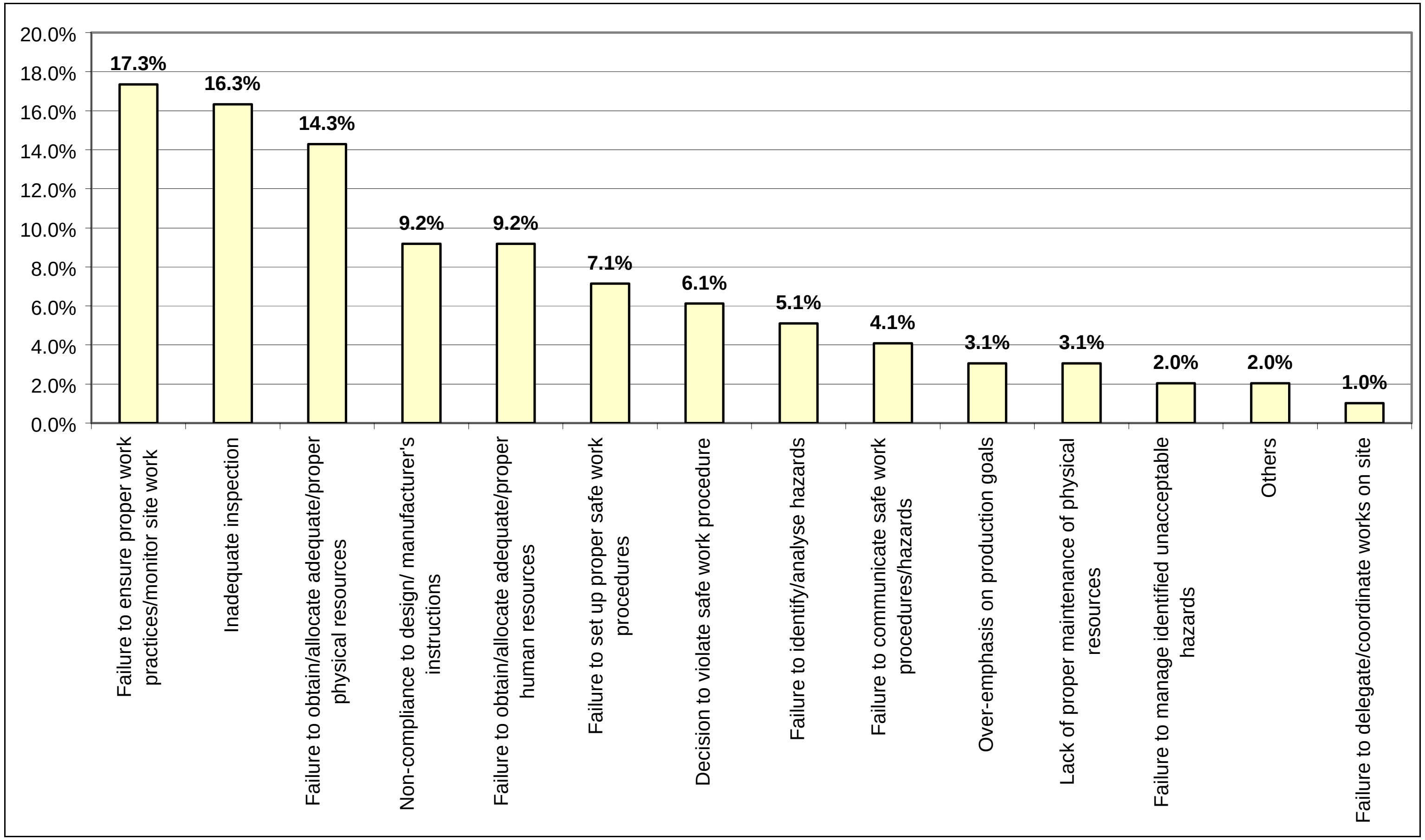


Figure A2.8 Distribution of type of job factors related to site management

Appendix 3

SEMANTIC NETWORKS

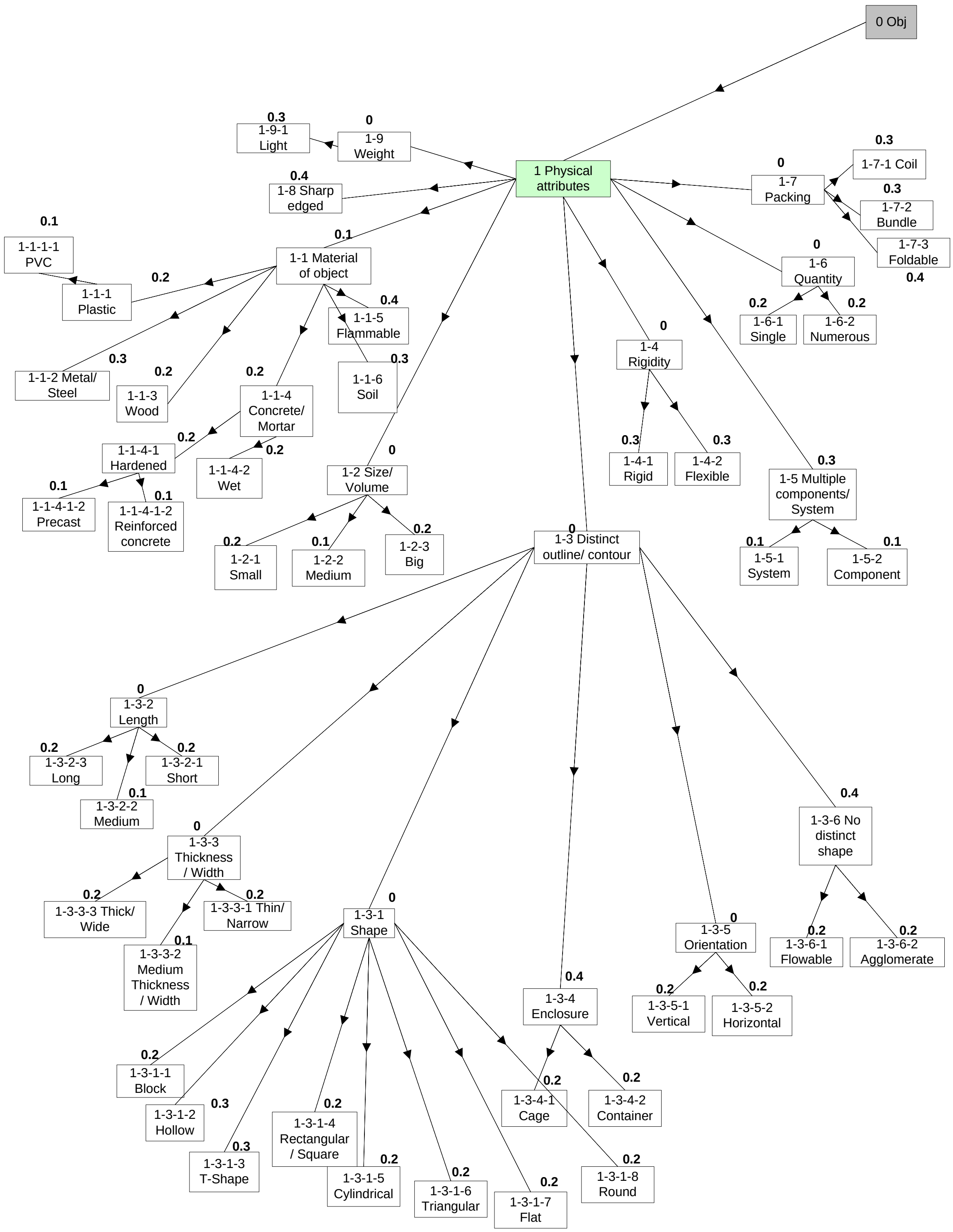


Figure A3.1 (a) Semantic network for situational variable "Objects" – sub-concepts related to physical attributes

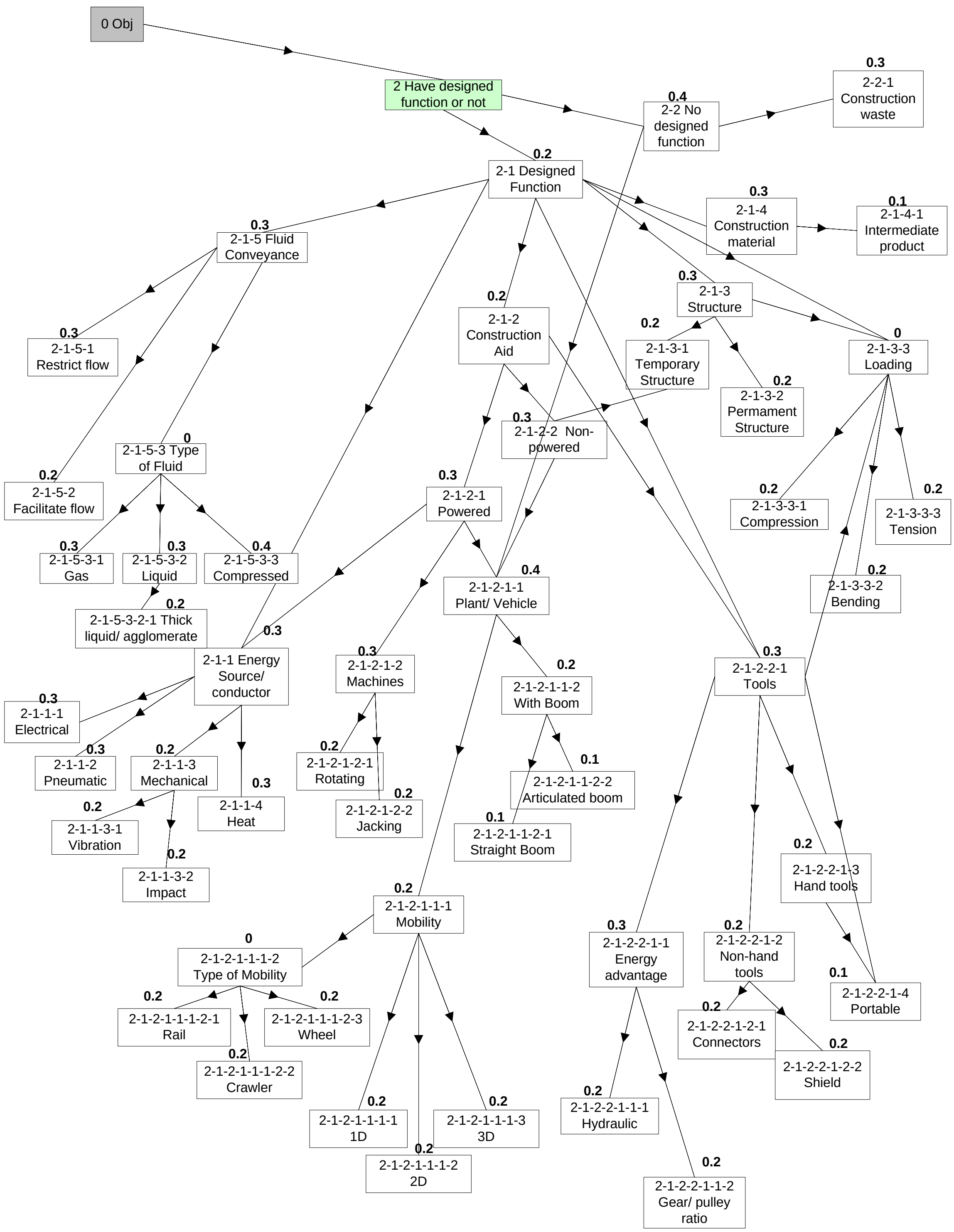


Figure A3.1 (b) Semantic network for situational variable "Objects" – sub-concepts related to functions

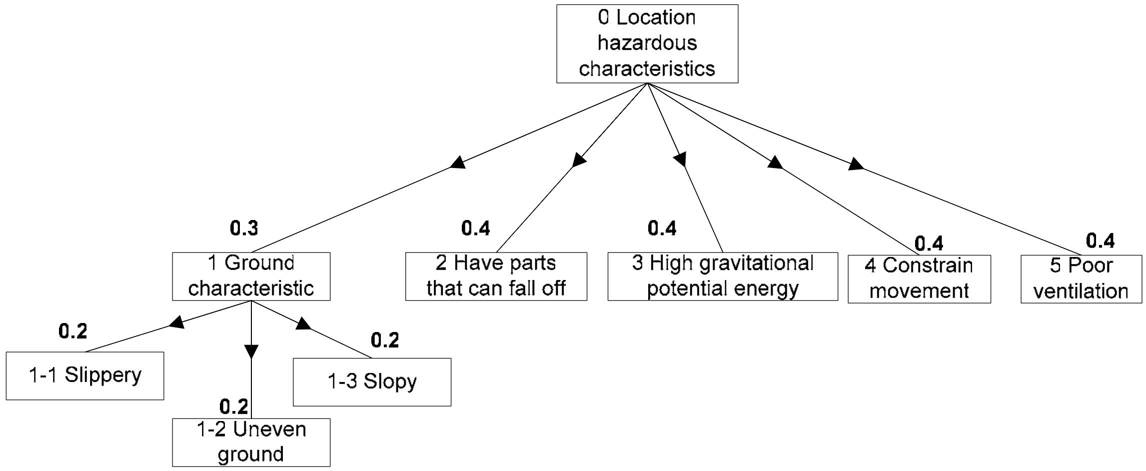


Figure A3.2 Taxonomy tree for situational variable “Location”

Appendix 4

VALIDATION OF THE POISSON DISTRIBUTION FOR CONSTRUCTION INCIDENTS

VALIDATING THE POISSON MODEL FOR CONSTRUCTION INCIDENTS

Data source

The data for this study has been obtained from the Safety Department of the Land Transport Authority (LTA) of Singapore. Since 1998, the LTA has implemented a computerised system called the Safety Information System (SITS) to capture information on incidents that occurred on LTA construction sites. The SITS contains incidents of all severity, from incidents with no injury to incidents involving fatalities, and LTA had been nurturing a transparent and non-penalising culture where reporting of these incidents is greatly encouraged. This approach had helped them to collect a relatively large amount of incident information.

Table A4.1 List of contracts chosen for analysis

Contract	Contract Description
A	Above ground construction
B	Above ground construction
C	Other underground construction work
D	Underground station construction
E	Underground station construction
F	Underground station construction
G	Underground station construction
H	Underground station construction with tunnelling work
I	Underground station construction with tunnelling work
J	Underground station construction with tunnelling work
K	Underground station construction with tunnelling work
L	Underground station construction with tunnelling work
M	Underground station construction with tunnelling work
N	Underground station construction with tunnelling work

In all, fourteen contracts with sufficient data points for statistical inference were chosen for the analysis as depicted in Table A4.1. All fourteen contracts are part of the Mass Rapid Transit (MRT) or Light Rail Transit (LRT) construction projects that either have been recently completed or are still on-going. Most of the contracts involve construction of railway stations, and the majority of the stations are underground. Besides construction of underground stations, Contracts H to N also include considerable tunnelling work. Unlike the other underground construction contracts, Contract C involves the construction and installation of railway components in the underground tunnels. The above ground construction contracts include construction of above ground stations, train depot for parking and maintenance of the trains, and viaducts for the railway system.

Furthermore, to minimise effects due to instability in the reporting and recording of incidents during the early stage of implementing SITS in the projects, initial data of about 100,000 man-hrs have been removed from each contract. This corresponds to between 4-6 months of the contracts. The data during this period has demonstrated exceptionally high variance or exceptionally low number of incidents, and would introduce unnecessary noise into the analysis, if included. These contracts are generally over 3.5 years long, involving several million man-hrs so that the data discarded represents only a small portion of the project.

Goodness-of-fit Test

The appropriateness of the Poisson distribution in modelling the random component of incident causation has been tested using the chi-square goodness-of-fit test (Conover, 1980; Bendell, 1991) and the dispersion test (Cox and Lewis, 1966; Nicholson, 1985; Nicholson and Wong, 1993). The former is one of the most commonly used tests to determine the goodness-of-fit of a distribution to some observed data, in which each data point is assumed to be an independent observation of the random variable $X(t)$. The statistic, T, for this test follows the chi-square distribution and is given by

$$T = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i} \qquad \textbf{(A4.1)}$$

where, O_i is the number of observed data in class *i* of the data (e.g., class *i* may be the class with x_i number of incidents in the time intervals observed), and E_i the expected number of observed data in that class as given by the Poisson distribution so that,

$$E_i = p_i \bullet \Sigma O_i \qquad \textbf{(A4.2)}$$

in which, p_i is the probability that $X(t) = x_i$, given by Eq. (1).

The classes for the test have been designed carefully to ensure that the assumptions of the test are not violated. For example, small values of the expected number E_i of observed data in class i can lead to poor match between the chi-square distribution and the actual distribution of T. This problem is resolved by applying a conservative rule of thumb proposed by Cochran (1954), in which the expected number of occurrences in each class must be greater than one, and more than 80% of the expected number of occurrences in all classes must be greater than 5. If the expected number of occurrences in a class is too low, the class is merged with adjacent classes to increase the E_i (Montgomery and Runger, 1999).

Dispersion Test

A key characteristic of the Poisson process is that the mean rate of arrival, λ, is equal to the standard deviation. As demonstrated by Bendell (1991), the chi-square test is not able to detect whether the sample's coefficient of variation (standard deviation/mean) is significantly different from unity. Instead, the dispersion test

(Nicholson, 1985; Nicholson and Wong, 1993; Cox and Lewis, 1966) has been utilized to validate this aspect of the model. The statistic, *H*, for this test also has a chi-square distribution and is given by

$$H = \sum_{i=1}^{c} \frac{(x_i - \bar{x})^2}{\bar{x}} \qquad \textbf{(A4.3)}$$

where x_i is the number of incident occurrences in the time interval *i*, and $\bar{x}$ the average number of incident occurrence in an interval, with *c* being the number of intervals in the sample.

In the analysis, the time interval size of 50,000 man-hours worked was chosen arbitrarily to ensure meaningful aggregation of incident occurrences. It is sufficiently small to prevent loss of information when incident counts are merged into large intervals. On the other hand, the intervals are not too small as to cause significance of errors contained in data to be amplified. In this regard, all the samples obtained for the contracts have more than adequate number of intervals and incidents necessary for a valid test. Specifically, the number of time intervals range from 44 to over 500 (see Table A4.2) and minimum number of incidents in the samples is 37, well exceeding the recommended minimum of 20 and 33, respectively (Nicholson and Wong, 1993).

DISCUSSION OF RESULTS

The results of the above tests are shown in Table A4.2 as *P*-values corresponding to the probabilities for the chi-square of the respective computed *T* and *H* statistics. Evident from Table A4.2, the data from all the contracts except for Contract A, have their *P*-values well exceeding 0.01, indicating that the distribution of the observed data corresponds well to that of a Poisson process. Closer analysis of the observed data in Contract A also shows that the Poisson process would be equally valid, as discussed shortly.

Table A4.2 Analysis results based on complete contract data

Contract	Dispersion test *P*-value	Chi-square test *P*-value	Mean arrival rate $\hat{\lambda}$	Coeff. of Variance	No. of Intervals of 50,000 mhr
A	*0.002*	*0.005*	0.209	1.190	532
B	0.529	0.948	0.154	0.989	188
C	0.719	0.635	0.455	0.892	66
D	0.055	0.396	0.551	1.239	98
E	0.238	0.437	0.795	1.145	44
F	0.562	0.792	0.295	0.972	105
G	0.058	0.101	0.748	1.216	115
H	0.169	0.445	0.605	1.093	210
I	0.016	0.476	0.851	1.243	175
J	0.056	0.012	0.416	1.209	125
K	0.065	0.100	1.258	1.238	89
L	0.043	0.594	1.024	1.279	85
M	0.337	0.882	0.689	1.053	103
N	0.034	0.288	1.255	1.284	94

The mean arrival rates, $\hat{\lambda}$, range from 0.154 to about 1.26 incidents per 50,000 man-hrs worked at the sites. This mean arrival rate is a parameter of the Poisson process that is dependent on the systematic factors contributed by the situational variables and the quality of SMS, as expressed in Figure 5-6 of Chapter 5. Generally, it can be observed from Table A4.3 that the contracts with only above ground construction (Contracts A and B) have the lowest incidents in contrast to the contracts with tunneling works, which would have greater exposure to risks having an average mean rate of 0.871 incidents per 50,000 man-hrs worked. On average, the contracts with underground station works alone are intermediate with an average mean rate of 0.597 incidents per 50,000 man-hrs worked. The dispersion of the mean rates within each category of works may be attributed to the other systematic factors, including SMS quality. A precise correlation of the mean arrival rates to the systematic factors warrants a more detailed study which is presently outside the scope of this paper.

Table A4.3 Contract descriptions based on ranked arrival rates

Contract	Mean arrival rate, $\hat{\lambda}$	Contract Description
B	0.154	Above ground construction
A	0.209	Above ground construction
F	0.295	Underground station construction
J	0.416	Underground station construction with tunnelling work
C	0.455	Other underground construction work
D	0.551	Underground station construction
H	0.605	Underground station construction with tunnelling work
M	0.689	Underground station construction with tunnelling work
G	0.748	Underground station construction
E	0.795	Underground station construction
I	0.851	Underground station construction with tunnelling work
L	1.024	Underground station construction with tunnelling work
N	1.255	Underground station construction with tunnelling work
K	1.258	Underground station construction with tunnelling work

With respect to Contract A, which failed the homogeneous Poisson distribution test above, an analysis of the incident rate over time shows that there was a significant reduction in the rate of incidence after interval 120 on Figure A4.1 depicting the number of incidents over time intervals of 50,000 man-hrs worked. Additional tests were performed by dividing the sample into 2 segments separating the difference. Three possible separation points were chosen to ensure that there are at least 34 incidents in each segment. These results are shown in Table A4.4, showing that the Poisson process, albeit a non-homogeneous one, is indeed valid as well. The average mean rate for the initial stages of the project was about 0.40 incidents per 50,000 man-hrs worked, compared to the significantly reduced mean rate of about

0.12 incidents per 50,000 man-hrs worked. This reduction could be attributed to the difference in the nature of work of the first part involving some basement construction or due to significant improvement in the SMS following the occurrence of earlier incidents.

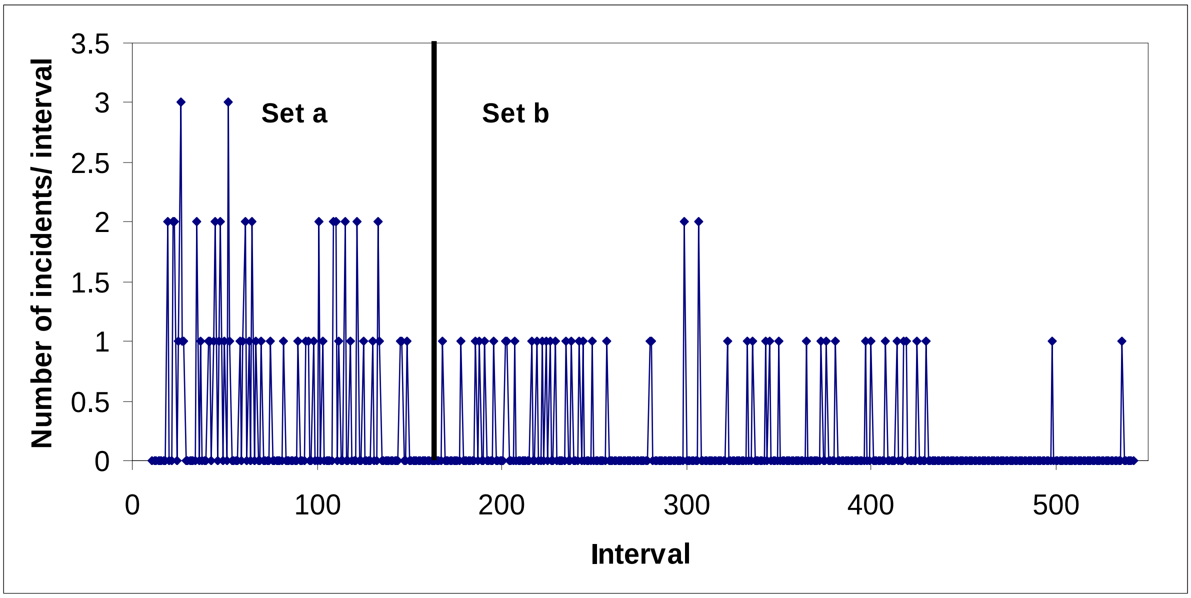


Figure A4.1 Time series plot of number of incidents per 50,000 man-hours for contract A

Table A4.4 Additional tests' results for contract A

Separation point (interval no.)	Sub-set	Dispersion test *P*-value	Chi-square test *P*-value	Mean arrival rate, $\hat{\lambda}$	Coeff. of Variance	No. of Intervals of 50,000 mhr
140	a	0.064	0.037	0.469	1.196	130
	b	0.794	0.678	0.124	0.782	402
180	a	0.025	0.021	0.388	1.225	170
	b	0.687	0.859	0.124	0.967	362
220	a	0.037	0.063	0.357	1.182	210
	b	0.477	0.955	0.112	1.002	322

Appendix 5

RISK ASSESSMENT TREE AFTER HAZARD IDENTIFICATION ADAPTATION

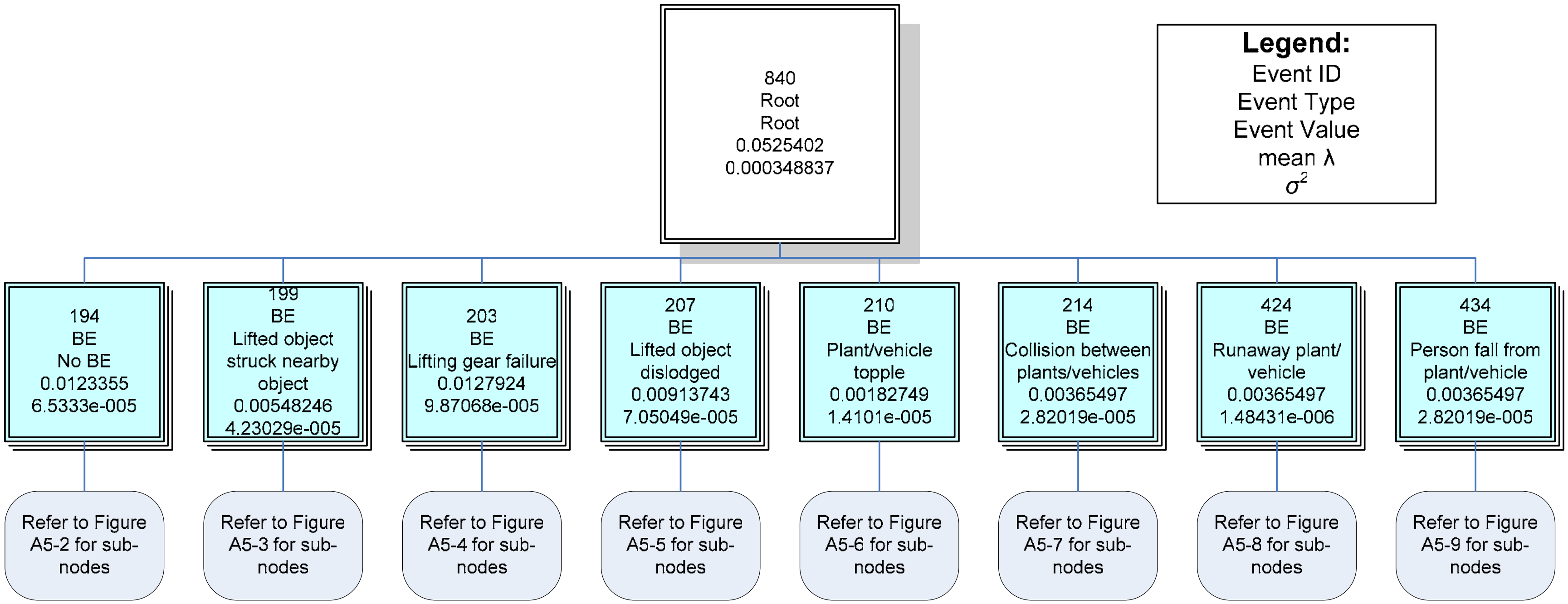


Figure A5.1 Risk assessment tree after hazard identification adaptation

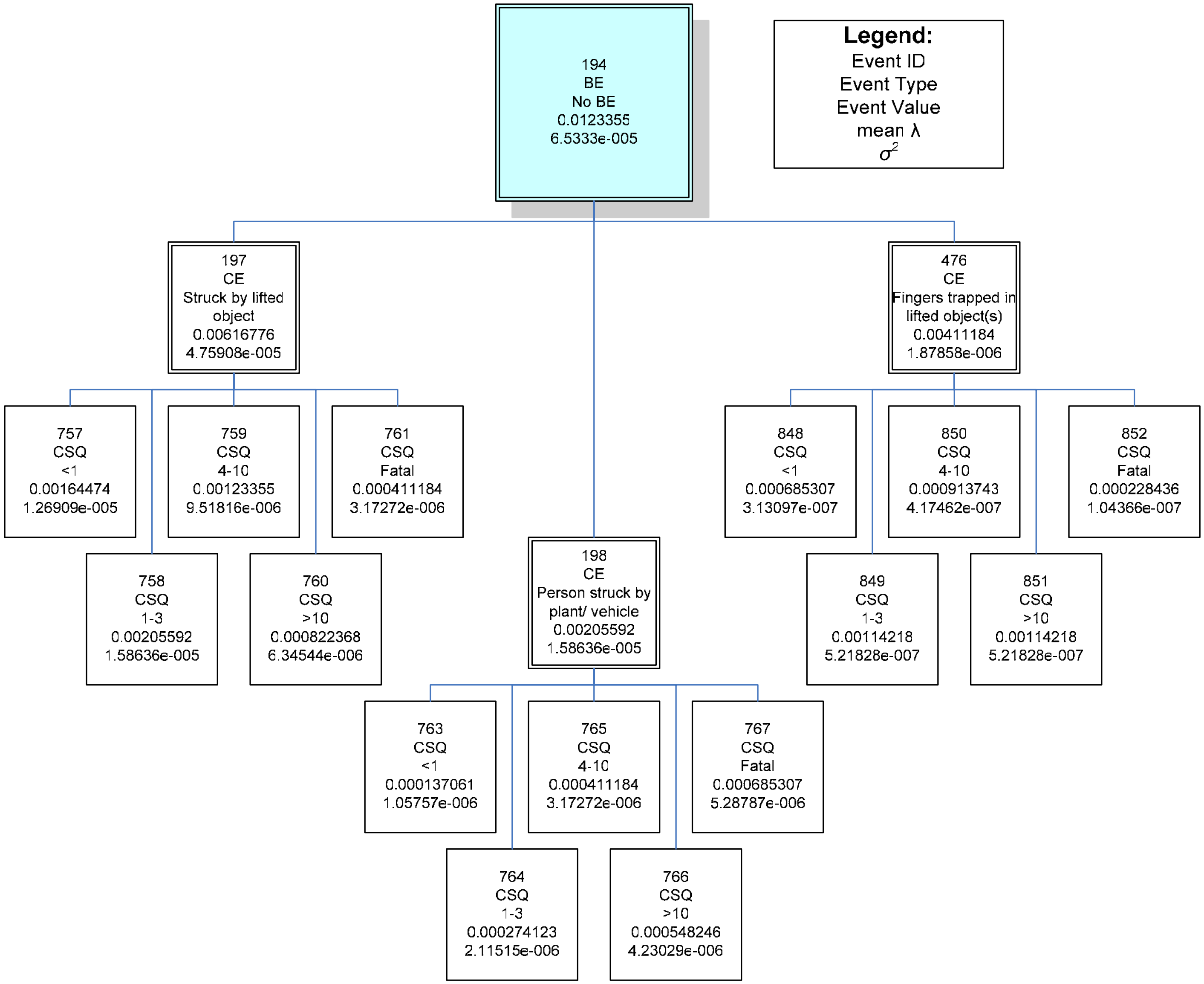


Figure A5.2 Incident events under breakdown event “No BE”

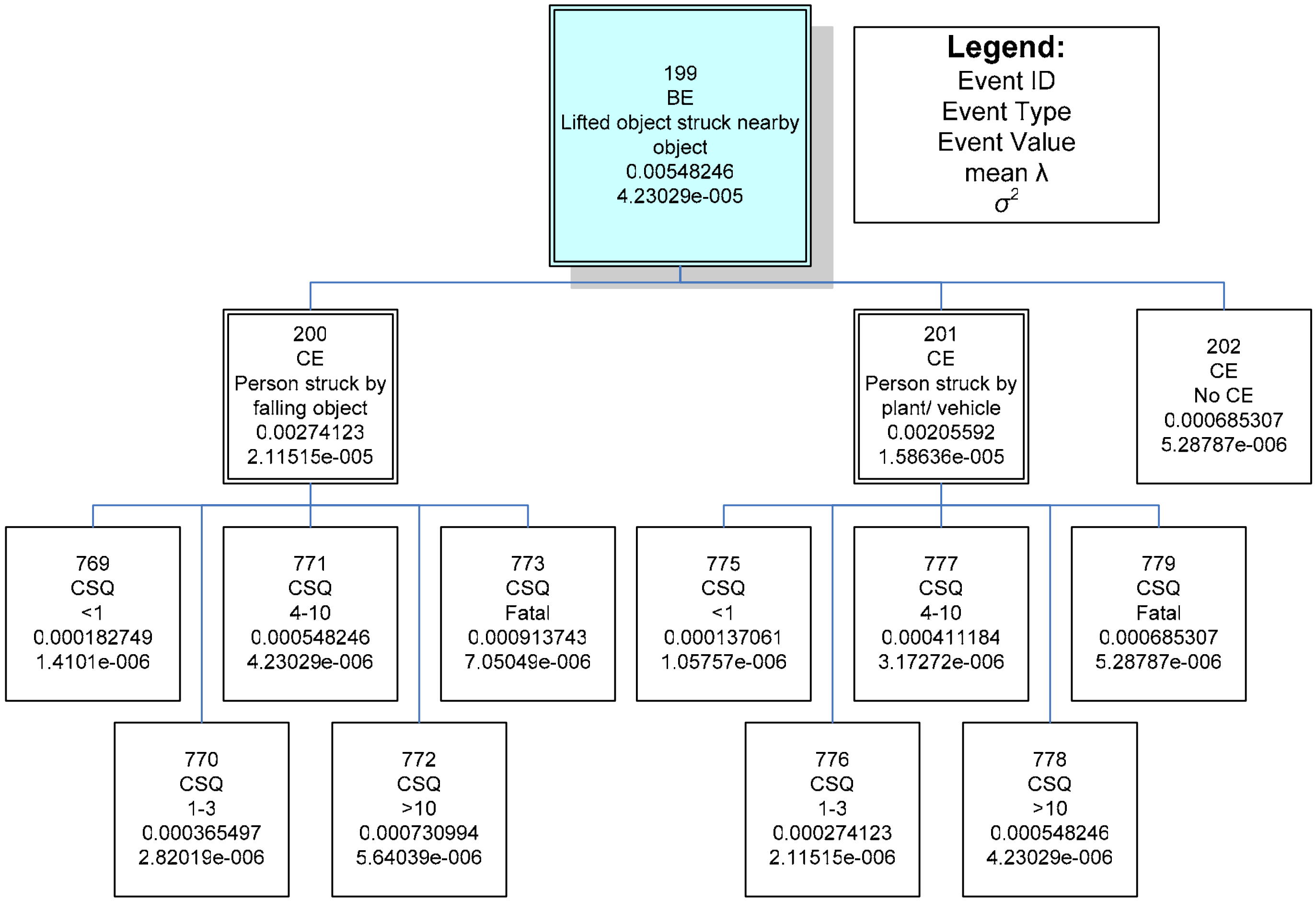


Figure A5.3 Incident events under breakdown event "Lifted object struck nearby object"

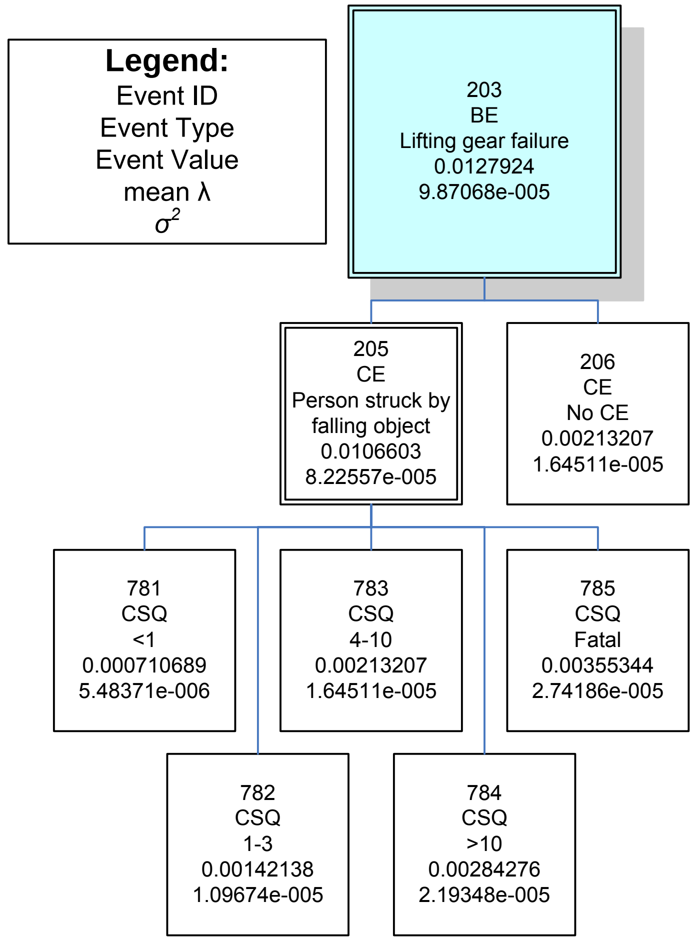


Figure A5.4 Incident events under breakdown event “Lifting gear failure”

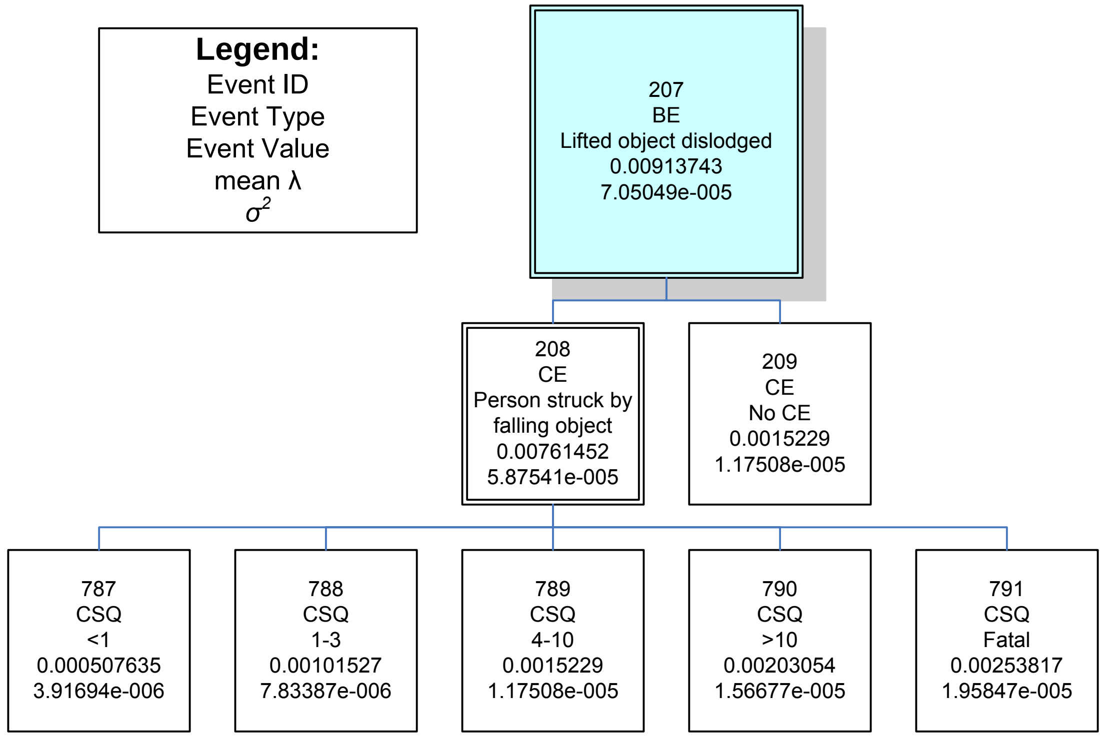


Figure A5.5 Incident events under breakdown event “Lifted object dislodged”

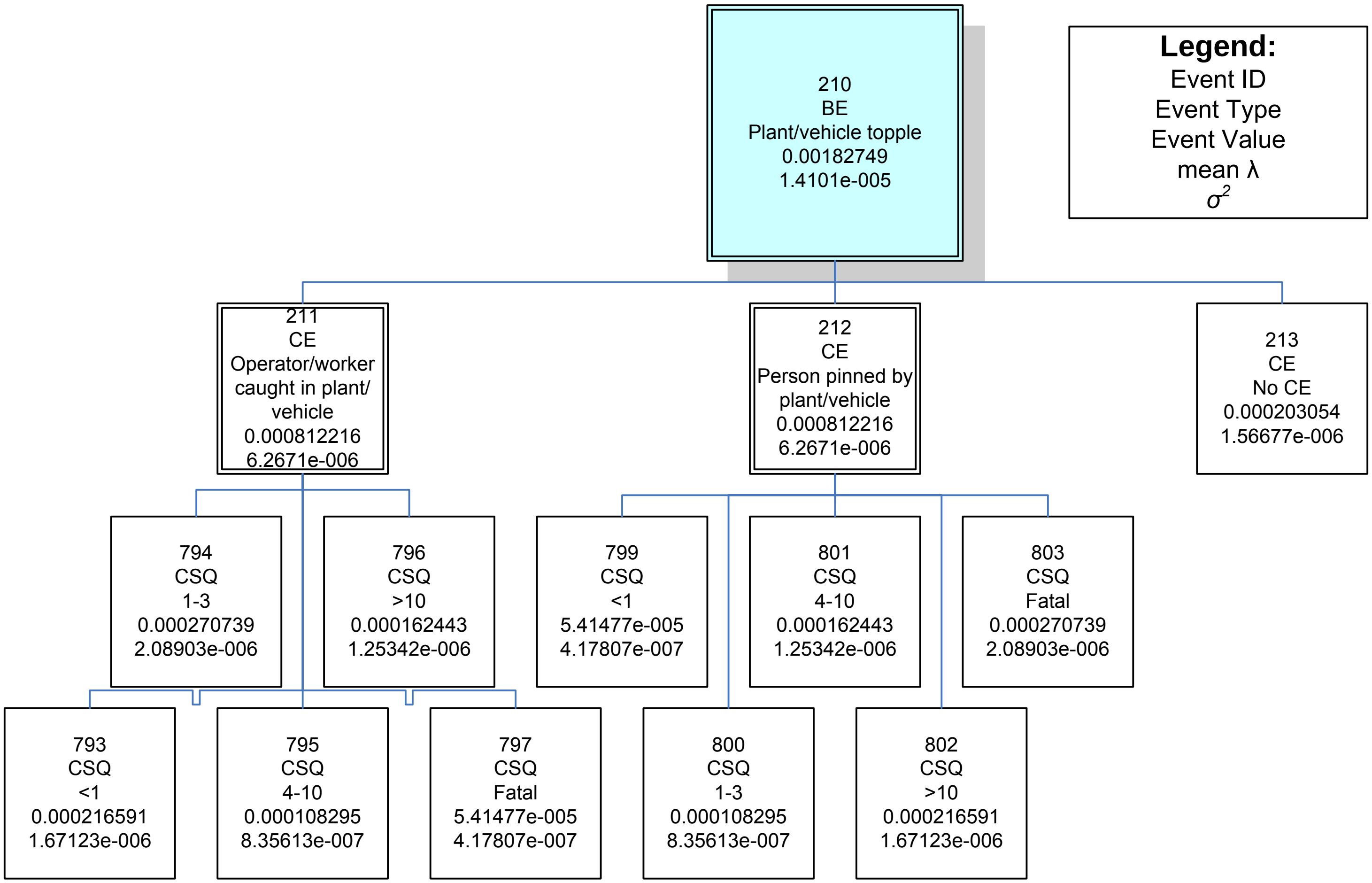


Figure A5.6 Incident events under breakdown event "Plant/vehicle topple"

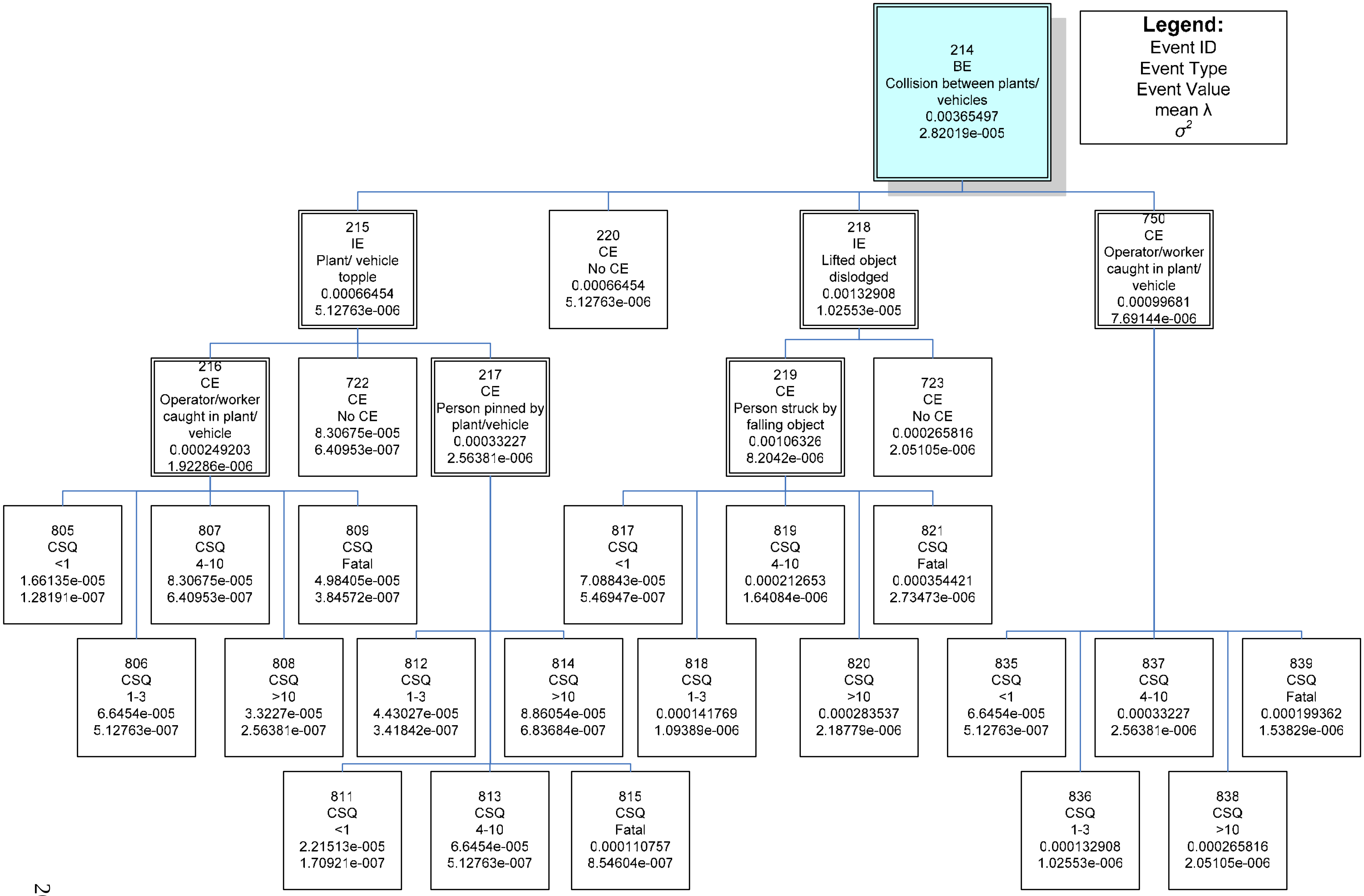


Figure A5.7 Incident events under breakdown event "Collision between plants/ vehicles"

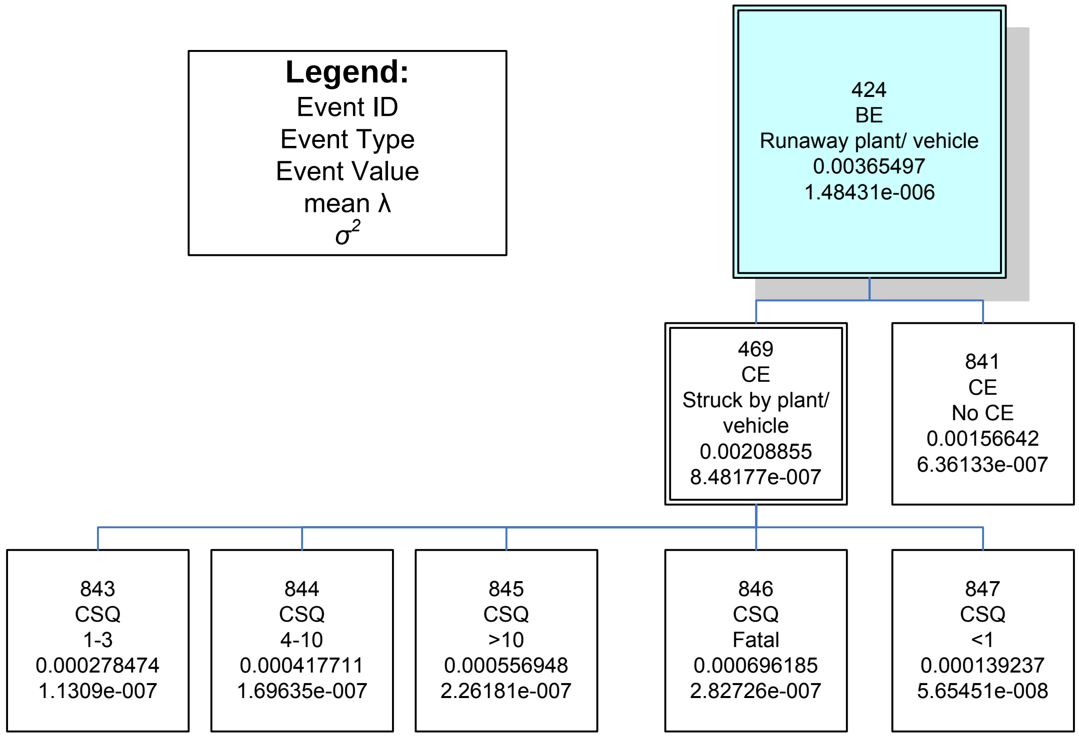


Figure A5.8 Incident events under breakdown event "Runaway plant/ vehicle"

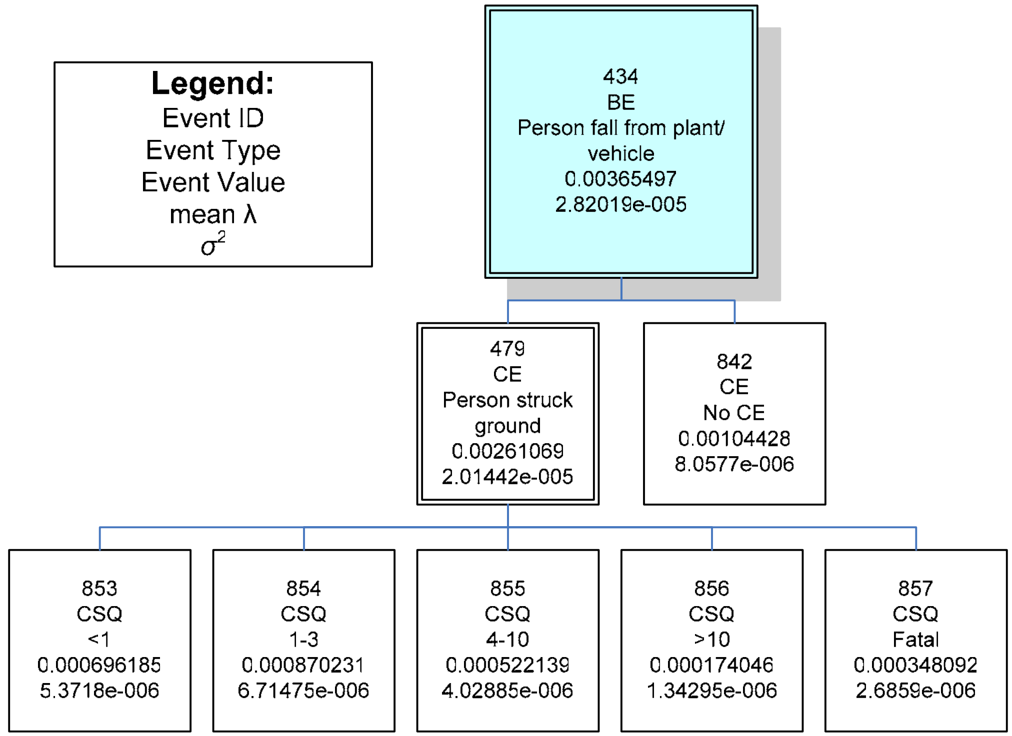


Figure A5.9 Incident events under breakdown event "Person fall from plant/vehicle"

Appendix 6

RESULTS OF BAYESIAN UPDATING

Table A6.1 Bayesian updating of leave nodes

Event Type	Pre-Event ID	Event ID	Event Value	*x*	*t*	$\bar{\lambda}'$	Initial Var(λ)	$\bar{\lambda}''$	Updated Var(λ)
CSQ	197	757	<1	0	278.966	0.001645	1.269E-05	0.0005217	1.277E-06
CSQ	197	758	1-3	0	278.966	0.002056	1.586E-05	0.0006522	1.596E-06
CSQ	197	759	4-10	0	278.966	0.001234	9.518E-06	0.0003913	9.577E-07
CSQ	197	760	>10	0	278.966	0.000822	6.345E-06	0.0002609	6.385E-07
CSQ	197	761	Fatal	0	278.966	0.000411	3.173E-06	0.0001304	3.192E-07
CSQ	198	763	<1	0	278.966	0.000137	1.058E-06	4.348E-05	1.064E-07
CSQ	198	764	1-3	0	278.966	0.000274	2.115E-06	8.695E-05	2.128E-07
CSQ	198	765	4-10	0	278.966	0.000411	3.173E-06	0.0001304	3.192E-07
CSQ	198	766	>10	0	278.966	0.000548	4.23E-06	0.0001739	4.257E-07
CSQ	198	767	Fatal	0	278.966	0.000685	5.288E-06	0.0002174	5.321E-07
CE	199	202	No CE	0	278.966	0.000685	5.288E-06	0.0002174	5.321E-07
CSQ	200	769	<1	0	278.966	0.000183	1.41E-06	5.797E-05	1.419E-07
CSQ	200	770	1-3	0	278.966	0.000365	2.82E-06	0.0001159	2.838E-07
CSQ	200	771	4-10	0	278.966	0.000548	4.23E-06	0.0001739	4.257E-07
CSQ	200	772	>10	0	278.966	0.000731	5.64E-06	0.0002319	5.675E-07
CSQ	200	773	Fatal	0	278.966	0.000914	7.05E-06	0.0002898	7.094E-07
CSQ	201	775	<1	0	278.966	0.000137	1.058E-06	4.348E-05	1.064E-07
CSQ	201	776	1-3	0	278.966	0.000274	2.115E-06	8.695E-05	2.128E-07
CSQ	201	777	4-10	0	278.966	0.000411	3.173E-06	0.0001304	3.192E-07
CSQ	201	778	>10	0	278.966	0.000548	4.23E-06	0.0001739	4.257E-07
CSQ	201	779	Fatal	0	278.966	0.000685	5.288E-06	0.0002174	5.321E-07
CE	203	206	No CE	**1**	278.966	0.002132	1.645E-05	0.0031239	7.646E-06
CSQ	205	781	<1	0	278.966	0.000711	5.484E-06	0.0002254	5.518E-07
CSQ	205	782	1-3	0	278.966	0.001421	1.097E-05	0.0004509	1.104E-06
CSQ	205	783	4-10	0	278.966	0.002132	1.645E-05	0.0006763	1.655E-06
CSQ	205	784	>10	0	278.966	0.002843	2.193E-05	0.0009017	2.207E-06
CSQ	205	785	Fatal	0	278.966	0.003553	2.742E-05	0.0011272	2.759E-06
CE	207	209	No CE	0	278.966	0.001523	1.175E-05	0.0004831	1.182E-06

Table A6.2 Bayesian updating of leave nodes (cont'd)

Event Type	Pre-Event ID	Event ID	Event Value	x	t	$\bar{\lambda}'$	Initial Var(λ)	$\bar{\lambda}''$	Updated Var(λ)
CSQ	208	787	<1	0	278.966	0.000508	3.917E-06	0.000161	3.941E-07
CSQ	208	788	1-3	0	278.966	0.001015	7.834E-06	0.0003221	7.882E-07
CSQ	208	789	4-10	0	278.966	0.001523	1.175E-05	0.0004831	1.182E-06
CSQ	208	790	>10	**1**	278.966	0.002031	1.567E-05	0.0030917	7.567E-06
CSQ	208	791	Fatal	0	278.966	0.002538	1.958E-05	0.0008051	1.971E-06
CE	210	213	No CE	0	278.966	0.000203	1.567E-06	6.441E-05	1.576E-07
CSQ	211	793	<1	0	278.966	0.000217	1.671E-06	6.87E-05	1.682E-07
CSQ	211	794	1-3	0	278.966	0.000271	2.089E-06	8.588E-05	2.102E-07
CSQ	211	795	4-10	0	278.966	0.000108	8.356E-07	3.435E-05	8.408E-08
CSQ	211	796	>10	0	278.966	0.000162	1.253E-06	5.153E-05	1.261E-07
CSQ	211	797	Fatal	0	278.966	5.41E-05	4.178E-07	1.718E-05	4.204E-08
CSQ	212	799	<1	0	278.966	5.41E-05	4.178E-07	1.718E-05	4.204E-08
CSQ	212	800	1-3	0	278.966	0.000108	8.356E-07	3.435E-05	8.408E-08
CSQ	212	801	4-10	0	278.966	0.000162	1.253E-06	5.153E-05	1.261E-07
CSQ	212	802	>10	0	278.966	0.000217	1.671E-06	6.87E-05	1.682E-07
CSQ	212	803	Fatal	0	278.966	0.000271	2.089E-06	8.588E-05	2.102E-07
CE	214	220	No CE	0	278.966	0.000665	5.128E-06	0.0002108	5.159E-07
CE	215	722	No CE	0	278.966	8.31E-05	6.41E-07	2.635E-05	6.449E-08
CSQ	216	805	<1	0	278.966	1.66E-05	1.282E-07	5.27E-06	1.29E-08
CSQ	216	806	1-3	0	278.966	6.65E-05	5.128E-07	2.108E-05	5.159E-08
CSQ	216	807	4-10	0	278.966	8.31E-05	6.41E-07	2.635E-05	6.449E-08
CSQ	216	808	>10	0	278.966	3.32E-05	2.564E-07	1.054E-05	2.58E-08
CSQ	216	809	Fatal	0	278.966	4.98E-05	3.846E-07	1.581E-05	3.87E-08
CSQ	217	811	<1	0	278.966	2.22E-05	1.709E-07	7.027E-06	1.72E-08
CSQ	217	812	1-3	0	278.966	4.43E-05	3.418E-07	1.405E-05	3.44E-08
CSQ	217	813	4-10	0	278.966	6.65E-05	5.128E-07	2.108E-05	5.159E-08
CSQ	217	814	>10	0	278.966	8.86E-05	6.837E-07	2.811E-05	6.879E-08
CSQ	217	815	Fatal	0	278.966	0.000111	8.546E-07	3.513E-05	8.599E-08

Table A6.3 Bayesian updating of leave nodes (cont'd)

Event Type	Pre-Event ID	Event ID	Event Value	*x*	*t*	$\bar{\lambda}'$	Initial Var(λ)	$\bar{\lambda}''$	Updated Var(λ)
CE	218	723	No CE	0	278.966	0.000266	2.051E-06	8.432E-05	2.064E-07
CSQ	219	817	<1	0	278.966	7.09E-05	5.469E-07	2.248E-05	5.503E-08
CSQ	219	818	1-3	0	278.966	0.000142	1.094E-06	4.497E-05	1.101E-07
CSQ	219	819	4-10	0	278.966	0.000213	1.641E-06	6.745E-05	1.651E-07
CSQ	219	820	>10	0	278.966	0.000284	2.188E-06	8.994E-05	2.201E-07
CSQ	219	821	Fatal	0	278.966	0.000354	2.735E-06	0.0001124	2.752E-07
CE	424	841	No CE	0	278.966	0.001566	6.361E-07	0.001407	5.133E-07
CE	434	842	No CE	0	278.966	0.001044	8.058E-06	0.0003313	8.108E-07
CSQ	469	847	<1	0	278.966	0.000139	5.655E-08	0.0001251	4.562E-08
CSQ	469	843	1-3	0	278.966	0.000278	1.131E-07	0.0002501	9.124E-08
CSQ	469	844	4-10	**1**	278.966	0.000418	1.696E-07	0.00074	2.699E-07
CSQ	469	845	>10	0	278.966	0.000557	2.262E-07	0.0005003	1.825E-07
CSQ	469	846	Fatal	0	278.966	0.000696	2.827E-07	0.0006253	2.281E-07
CSQ	476	848	<1	0	278.966	0.000685	3.131E-07	0.0006078	2.463E-07
CSQ	476	849	1-3	**1**	278.966	0.001142	5.218E-07	0.0014183	5.747E-07
CSQ	476	850	4-10	0	278.966	0.000914	4.175E-07	0.0008104	3.284E-07
CSQ	476	851	>10	0	278.966	0.001142	5.218E-07	0.0010131	4.105E-07
CSQ	476	852	Fatal	0	278.966	0.000228	1.044E-07	0.0002026	8.21E-08
CSQ	479	853	<1	0	278.966	0.000696	5.372E-06	0.0002208	5.405E-07
CSQ	479	854	1-3	0	278.966	0.00087	6.715E-06	0.000276	6.756E-07
CSQ	479	855	4-10	0	278.966	0.000522	4.029E-06	0.0001656	4.054E-07
CSQ	479	856	>10	**1**	278.966	0.000174	1.343E-06	0.0025028	6.126E-06
CSQ	479	857	Fatal	0	278.966	0.000348	2.686E-06	0.0001104	2.703E-07
CSQ	750	835	<1	0	278.966	6.65E-05	5.128E-07	2.108E-05	5.159E-08
CSQ	750	836	1-3	0	278.966	0.000133	1.026E-06	4.216E-05	1.032E-07
CSQ	750	837	4-10	0	278.966	0.000332	2.564E-06	0.0001054	2.58E-07
CSQ	750	838	>10	0	278.966	0.000266	2.051E-06	8.432E-05	2.064E-07
CSQ	750	839	Fatal	0	278.966	0.000199	1.538E-06	6.324E-05	1.548E-07

Table A6.4 Initial frequency, variance, severity and risk values of incident events

Event ID	Pre-Event ID	Event Type	Event Value	$\overline{\lambda}'$	Initial Var(λ)	Initial *E*(Sev)	Initial Risk
840	Root	Root	Root	0.05254	3.49E-04	26.6835	1.401957
194	840	BE	No BE	0.01234	6.53E-05	16.9296	0.208836
199	840	BE	Lifted object struck nearby object	0.00548	4.23E-05	35.3208	0.193645
203	840	BE	Lifting gear failure	0.01279	9.87E-05	33.6389	0.430322
207	840	BE	Lifted object dislodged	0.00914	7.05E-05	33.6389	0.307373
210	840	BE	Plant/vehicle topple	0.00183	1.41E-05	23.4519	4.29E-02
214	840	BE	Collision between plants/vehicles	0.00365	2.82E-05	24.7832	9.06E-02
424	840	BE	Runaway plant/ vehicle	0.00365	1.48E-06	23.0667	8.43E-02
434	840	BE	Person fall from plant/vehicle	0.00365	2.82E-05	12.0476	4.40E-02
479	434	CE	Person struck ground	0.00261	2.01E-05	16.8667	4.40E-02
469	424	CE	Struck by plant/ vehicle	0.00209	8.48E-07	40.3667	8.43E-02
219	218	CE	Person struck by falling object	0.00106	8.20E-06	40.3667	4.29E-02
216	215	CE	Operator/worker caught in plant/vehicle	0.00025	1.92E-06	25.5667	6.37E-03
217	215	CE	Person pinned by plant/vehicle	0.00033	2.56E-06	40.3667	1.34E-02
215	214	IE	Plant/ vehicle topple	0.00066	5.13E-06	29.7708	1.98E-02
218	214	IE	Lifted object dislodged	0.00133	1.03E-05	32.2933	4.29E-02
750	214	CE	Operator/worker caught in plant/vehicle	0.001	7.69E-06	27.9667	2.79E-02
211	210	CE	Operator/worker caught in plant/vehicle	0.00081	6.27E-06	12.4	1.01E-02
212	210	CE	Person pinned by plant/vehicle	0.00081	6.27E-06	40.3667	3.28E-02
208	207	CE	Person struck by falling object	0.00761	5.88E-05	40.3667	0.307373
205	203	CE	Person struck by falling object	0.01066	8.23E-05	40.3667	0.430322
200	199	CE	Person struck by falling object	0.00274	2.12E-05	40.3667	0.110654
201	199	CE	Person struck by plant/ vehicle	0.00206	1.59E-05	40.3667	0.082991
197	194	CE	Struck by lifted object	0.00617	4.76E-05	11.5333	7.11E-02
198	194	CE	Person struck by plant/ vehicle	0.00206	1.59E-05	40.3667	0.082991
476	194	CE	Fingers trapped in lifted object(s)	0.00411	1.88E-06	13.3056	5.47E-02

Table A6.5 Updated frequency, variance, severity values and risk of incident events

Event ID	Pre-Event ID	Event Type	Event Value	$\bar{\lambda}''$	Updated Var(λ)	Updated *E*(Sev)	Updated Risk
840	Root	Root	Root	0.02925	5.57E-05	21.4098	0.626135
194	840	BE	No BE	0.00666	8.03E-06	14.7467	0.098226
199	840	BE	Lifted object struck nearby object	0.00174	4.26E-06	35.3208	0.061426
203	840	BE	Lifting gear failure	0.00651	1.59E-05	20.9827	0.136501
207	840	BE	Lifted object dislodged	0.00535	1.31E-05	27.3946	0.146453
210	840	BE	Plant/vehicle topple	0.00058	1.42E-06	23.4519	1.36E-02
214	840	BE	Collision between plants/vehicles	0.00116	2.84E-06	24.7832	2.87E-02
424	840	BE	Runaway plant/ vehicle	0.00365	1.33E-06	21.46	7.83E-02
434	840	BE	Person fall from plant/vehicle	0.00361	8.83E-06	17.4439	6.29E-02
479	434	CE	Person struck ground	0.00328	8.02E-06	19.2079	6.29E-02
469	424	CE	Struck by plant/ vehicle	0.00224	8.17E-07	34.9349	7.83E-02
219	218	CE	Person struck by falling object	0.00034	8.26E-07	40.3667	1.36E-02
216	215	CE	Operator/worker caught in plant/vehicle	7.9E-05	1.93E-07	25.5667	2.02E-03
217	215	CE	Person pinned by plant/vehicle	0.00011	2.58E-07	40.3667	4.25E-03
215	214	IE	Plant/ vehicle topple	0.00021	5.16E-07	29.7708	6.28E-03
218	214	IE	Lifted object dislodged	0.00042	1.03E-06	32.2933	1.36E-02
750	214	CE	Operator/worker caught in plant/vehicle	0.00032	7.74E-07	27.9667	8.84E-03
211	210	CE	Operator/worker caught in plant/vehicle	0.00026	6.31E-07	12.4	3.19E-03
212	210	CE	Person pinned by plant/vehicle	0.00026	6.31E-07	40.3667	1.04E-02
208	207	CE	Person struck by falling object	0.00486	1.19E-05	30.1159	0.146453
205	203	CE	Person struck by falling object	0.00338	8.28E-06	40.3667	0.136501
200	199	CE	Person struck by falling object	0.00087	2.13E-06	40.3667	0.0351
201	199	CE	Person struck by plant/ vehicle	0.00065	1.60E-06	40.3667	0.026325
197	194	CE	Struck by lifted object	0.00196	4.79E-06	11.5333	2.26E-02
198	194	CE	Person struck by plant/ vehicle	0.00065	1.60E-06	40.3667	0.026325
476	194	CE	Fingers trapped in lifted object(s)	0.00405	1.64E-06	12.175	4.93E-02

Table A6.6 Change (posterior – prior) in frequency, variance, severity and risk values of incident events

Event ID	Pre-Event ID	Event Type	Event Value	Change (Posterior – Prior Estimates)			
				$\bar{\lambda}$	Var(λ)	*E*(Sev)	Risk
840	Root	Root	Root	-0.02329	-2.93E-04	-5.2737	-0.77582
194	840	BE	No BE	-0.00567	-5.73E-05	-2.1829	-0.11061
199	840	BE	Lifted object struck nearby object	-0.00374	-3.80E-05	0	-0.13222
203	840	BE	Lifting gear failure	-0.00629	-8.28E-05	-12.656	-0.29382
207	840	BE	Lifted object dislodged	-0.00379	-5.74E-05	-6.2443	-0.16092
210	840	BE	Plant/vehicle topple	-0.00125	-1.27E-05	0	-2.93E-02
214	840	BE	Collision between plants/vehicles	-0.0025	-2.54E-05	0	-6.18E-02
424	840	BE	Runaway plant/ vehicle	-7.2E-06	-1.54E-07	-1.6067	-6.03E-03
434	840	BE	Person fall from plant/vehicle	-4.8E-05	-1.94E-05	5.39626	1.89E-02
479	434	CE	Person struck ground	0.00067	-1.21E-05	2.3412	1.89E-02
469	424	CE	Struck by plant/ vehicle	0.00015	-3.08E-08	-5.4318	-6.03E-03
219	218	CE	Person struck by falling object	-0.00073	-7.38E-06	0	-2.93E-02
216	215	CE	Operator/worker caught in plant/vehicle	-0.00017	-1.73E-06	0	-4.35E-03
217	215	CE	Person pinned by plant/vehicle	-0.00023	-2.31E-06	0	-9.16E-03
215	214	IE	Plant/ vehicle topple	-0.00045	-4.61E-06	0	-1.35E-02
218	214	IE	Lifted object dislodged	-0.00091	-9.22E-06	0	-2.93E-02
750	214	CE	Operator/worker caught in plant/vehicle	-0.00068	-6.92E-06	0	-1.90E-02
211	210	CE	Operator/worker caught in plant/vehicle	-0.00055	-5.64E-06	0	-6.88E-03
212	210	CE	Person pinned by plant/vehicle	-0.00055	-5.64E-06	0	-2.24E-02
208	207	CE	Person struck by falling object	-0.00275	-4.69E-05	-10.251	-0.16092
205	203	CE	Person struck by falling object	-0.00728	-7.40E-05	0	-0.29382
200	199	CE	Person struck by falling object	-0.00187	-1.90E-05	0	-0.07555
201	199	CE	Person struck by plant/ vehicle	-0.0014	-1.43E-05	0	-0.05667
197	194	CE	Struck by lifted object	-0.00421	-4.28E-05	0	-4.86E-02
198	194	CE	Person struck by plant/ vehicle	-0.0014	-1.43E-05	0	-0.05667
476	194	CE	Fingers trapped in lifted object(s)	-6E-05	-2.37E-07	-1.1306	-5.37E-03