A Framework of Occupational Safety Hazard Analysis and Control to Support Enterprise Health and Safety Information Digitalization

by

Changcui Qiu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Engineering Management

Department of Mechanical Engineering
University of Alberta

ABSTRACT

Occupational safety hazards and risk management in construction industry are major worldwide concerns due to its unique dynamic nature of the working environment. Safety related documentations such as industry injury reports, physical demands analysis (PDA) and standard operating procedures (SOP) contain valuable information to support the risk control and prevention algorithm. Systematic analysis of the risk factors on cause-and effect relationship and safety attributes of the incident injury reports can make significant contributions to the construction industry and is an ideal approach for the occupational performance and risk evaluation. This thesis presents a blended study on occupational safety hazard control and risk assessment to support the enterprise health and safety digital information digitalization. The study first explores a strategy of attribute-based identification and degree of risk classification by applying distinctive quantitative analysis on level of injuries. Such analysis is conducted with the support of both physical demands analysis (PDA), standard operating procedures (SOP) and a large number of incident reports. To further expose the causality links and relation analysis of potentially hazardous activities and construction hazards, data mining analysis is then carried out following the attribute-based risk assessment model by using the RapidMiner. Then, a conceptual digitalization framework is also proposed to assess digital information mapping and future predictive measurements for effective risk management and up-to-date risk factor evaluation. The study discussed how the digitalization framework is essential to the construction industry by exposing the common risks that the industry is currently facing. Finally, a health and safety (H&S) information flowchart, as well as an information management and control framework for the H&S department is also proposed. The blended analysis is expected to help the construction industry

identify relationships on causes of safety hazards, key safety attributes and ergonomic characteristics, as well as the level of injury and their corresponding risk controls and preventions. The frameworks support H&S information digitalization and are expected to facilitate H&S information management to improve workplace productivity and efficiency.

PREFACE

This thesis is an original work by Changcui Qiu and follows a monograph format. One journal paper and one conference paper related to this thesis have been published as listed below.

- **Qiu, C.**, Li, X. (2021). Blended Analysis of Occupational Safety Hazards and Risk Assessment Approach in the Construction Industry. In: et al. *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*. CSCE 2021. Lecture Notes in Civil Engineering, vol 251. Springer, Singapore. https://doi.org/10.1007/978-981-19-1029-6_38 Dr. Xinming Li was the supervisory authority and was involved with concept formation and manuscript composition.
- **Qiu, C**. and Li, X. (2022). Blended analysis of Occupational Safety Hazards and Digital Transformation of Risk Assessment in the Construction Industries. *Canadian Journal of Civil Engineering.* CJCE 2022. Special Issue on Automation and Digital Transformation in the Construction Industry. http://dx.doi.org/10.1139/cjce-2022-0036 Dr. Xinming Li was the supervisory authority and was involved with concept formation and manuscript composition.

ACKNOWLEDGEMENTS

I would like to acknowledge those individuals who accompanied me on this special journey. Only with the help of them was I able to embrace the challenges that awaited.

I would like to first express my sincere gratitude to my supervisor Dr. Xinming Li, for accepting me as a master student in her research team, this journey has been a time of immense growth and immersive experiences. I would like to thank her for her valuable guidance, passion, kindness, and ongoing support in the entire research process; and especially, I want to express my appreciation to her for guiding me through the detail required in such a cheerful and helpful manner. I am particularly thankful for all of the prompt emails she sent, her immense engineering knowledge and insights, her availability for regular updates and her sincere smiles for every online meeting.

I would also like to extend my gratitude to Dr. Haitao Yu and Dr. Mohammed Sadiq Altaf for sharing their knowledge, advice, and wisdom over the course of the enterprise digital twin project with the local collaborative industry partner. I am sincerely grateful for their time commitment, feedbacks, opportunities, and grant supports. Thank you for taking time to meet me on campus during the weekends and thank you for always responding to my updated project related emails and meeting requests. I am also extremely thankful to my graduate research fellow, Aswin Ramaswamy Govindan. Thank you for sharing all the industry insights, brilliant ideas, and thoughts, and always asking thoughtful questions and keeping me posted. My educational endeavours would not have been this meaningful without the support of him on our digital twin project.

My warmest regards and gratitude go to my dedicated family and dearest friends. I have been blessed to have such a supportive dad who continuously provided moral and spiritual supports throughout my life with his love and his greatest patience. And to my mom, thank you for always praying for me whenever I encountered difficulties and emotional breakdowns. Thank you for my lovely family members and friends who were always keen to know how my research work was proceeding and what I was studying.

Finally, but by no means least, I want to say special thank you to many of my good friends, supportive stuff and dearest professors and instructors who supported me in many directions and advised me throughout my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CSHM	Construction Safety and Health Monitoring
EMIS	Enterprise Management Information System
FE	Frequency of Exposure
H&S	Health and Safety
IP	Incident Probability
MIS	Management Information System
NIOSH	National Institute for Occupational Safety and Health
OHS	Occupational Health and Safety
OSHA	Occupational Safety and Health Administration
PC	Potential Consequence
PDA	Physical Demands Analysis
PPE	Personal Protective Equipment
SOP	Standard Operating Procedures
SMS	Safety Management System
WMSDS	Work-related Musculoskeletal Disorders

CHAPTER 1: INTRODUCTION

1.1 Background and Motivation

Given the complex existence of the construction working environment in most construction industries, the study of construction hazards and their associated risk assessment is limited due to their regulatory-based and reactive nature (Hallowell and Gambatese 2007; Uzo and Mohamed 2018). Workers were unaware of the risks if they were effective in performing jobs, and each underlying imminent risk of sub-standard work practice tends to move them into a state of individual complacency that can cause more injuries. The construction industry has been implicated in an increasing number of significant occupational impacts on workers' unsafe behaviours, undesirable safety events and conditions (McCabe et al. 2008; Fang et al. 2015), and specific characteristics inherent to H&S risks tend to create additional barriers that increase the probability of hazardous behaviour occurrence and ergonomic risk for this industry. Systematic and statistical analysis of occupational risk and safety hazards is critical to avoid construction accidents. Meaningful data with correlation can be extracted from a number of industry injury reports, physical demands analysis (PDA) and standard operating procedures (SOP) to generate useful results. The risk evaluation and potential correlation inferred from the recent studies contribute to effective risk identifications and help identify potentially risky activities and construction hazards. Quantitative evaluation and method are often more feasible, and arguably more reliable when considering a large number of exposures and accidents. Thus, having the ability to quantitatively examine occupational performance by utilizing industry injury reports is key from both practical and economical perspectives.

One prominent way to investigate the occupational risk and hazard interrelationship is through the attribute-based approach. This powerful approach allows the baseline characteristics of construction risk attribute to be uniquely defined and any construction incident cases can be interpreted as the resulting outcome based on the occurrence of significant safety risk attributes (Prades 2014). The attribute-based risk model for measuring safety risk was first introduced in 2012 (Esmaeili and Hallowell 2012) and has become widely popular in recent research (Prades 2014). Considering the limitation that existing model only discusses the quantified interactions between various exposure frequencies and their relationship to the degree of probable consequence, this thesis focuses on a concise risk quantification based on the frequency of exposure, incident probability and potential consequence for the degree of risk assessment. A blended analysis is first introduced based on an attribute-based risk approach together with data mining analysis to determine the level of injury and their corresponding risk controls using potential linkage of risky activities. Limitations on risk analysis for distinct construction related job tasks may still exist in utilizing the content generated in injury reports and can be improved by using both PDA and SOP for providing data on station-by-station injury exposure, physical demand and other environmental constraints (Li et al. 2019). Data mining analysis will further explore the associations and potential linkage of potentially risky activities. Results from this blended analysis deliver an overview of the potential risk factors causing unforeseen injuries and illnesses to identify specific tasks and job titles that are linked to these risk factors. It could provide further knowledge of causality links of risk factors and how to mitigate injuries and illnesses.

In addition, the emergence of digitalization in the construction industry enables real-time injury data assessments. Therefore, the complex industry could benefit from the robust safety information system and digitalization deployment method for better overall H&S performance. Grieves (2015) considers data from the "Physical" to the "Virtual" to be raw and requires processing, these important data can store knowledge across digital models with advanced degrees of interpretation. Aside from enhancing digital H&S data knowledge and hazard control, such information-intensive digitalization and associated technologies play a decisive role in diminishing potential hazard impact and defining control strategies. A conceptual framework based on the developed blended analysis model with a conceptual optimization algorithm, which is also proposed allows future risk prediction and reliable prioritizing of identified risks by inquiring objective judgments. In this thesis, the proposed framework is intended to integrate the blended analysis into a digital platform that ultimately links to a digital twin-driven design where digital mapping of physical data generated in the use of safety related data can be assessed and evaluated for future risk identification. Integrating safety related documents into occupational safety hazards assessment and risk management with digitalization is expected to provide a decision tool for injury prevention.

The industry of construction is known for its complexity, with a wide range of interdependent activities occurring concurrently. Information managed by the H&S division is massive, especially when it comes to information transferring and communication among various departments when human manual information processing is required. Due to the lack of human resources and expert professional skills, traditional strategic management in the civil engineering construction process is mostly manual supervision, which inherently affects the quality of safety and health information management (Yin 2019). Neglect of safety regulations and communications technology, lack of data supervision, resulting in accidents, injuries, loss of professional work capacity, chronic diseases, and, as a result, a reduction in the economic efficiency of construction industry organizations. In addition, each management system has primary focuses on its own and is used by different employees, which barriers information transmission and results in information loss. This would eventually result in not truly displaying the reality of H&S data to all personnel and increasing the chance of getting errors. Thus, the seamless and automated connection among the digital platform is critical for information integration. Within this objective, the systematic control of the information among different departments flow will be identified by demonstrating the reasoning, restrictions, function ability of the flow control in an efficient way. Efforts should be made in this regard to promote the integration of H&S information management that specifically focus on the H&S department, thereby effectively ensuring and enriching the continuous improvement and rationalisation of the construction database. The study also proposed both H&S department information flow diagram and departmental information management and control framework to create a comprehensive digital representation for high-level H&S management. A strategic management view that links occupational health and safety (OHS) concerns to the organization's mission can establish a context for expansion, profitability, and performance. This study emphasised that common information flow obstacles that occurred during the digital adoption in the construction industry, such as misinformation and misunderstanding of the flow, can be overcome through systematic digital solutions by enhancing information management integration of H&S with cross-departmental functions.

1.2 Research Objectives

With this background and motivation in mind, the following research questions are used in this study to examine and answer the research problems.

1) How do we efficiently identify and assess health and safety risks given in the context of the complex existence of the construction working environment?

2) How do we accurately assess, retrieve, and utilize the safety-related documents and analyze real-time injury data assessments through digitalization?

3) How do we effectively manage and adopt key technologies in information management and control when the scope and complexity of a health and safety management system varies?

To fulfill the above defined research questions and problems, the following objectives are being pursued by this research:

Objective 1: Develop a blended analysis on occupational hazards assessment and risk identification through both attribute-based and data mining approaches.

The first objective is served to capture the dynamics of construction work and address the challenges in existing risk assessment methods, including (1) lack of practical safety risk quantification method; (2) insufficient levels of detail on data sources; and (3) lack of focus on interactions among risk factors. The objective will identify the significant factors which affect injury severity involved in the construction projects, the level of injury and their corresponding risk controls can be defined using potential linkage of risky activities. This objective demonstrates a comprehensive knowledge baseline of safety risk quantification with the support of safety-related documentation. The developed blended analysis can reveal causality analysis of potentially hazardous activities and identify construction risks with controls and preventions.

Objective 2: Propose a conceptual digitalization framework using both safety related documents and developed blended analysis for proactive digital H&S information mapping and up-to-date risk factor evaluation.

The motivation underlying this objective is to address the need for an adapted information-intensive framework for digitalization. Due to the high time, labor related to manual analysis, most part of the valuable safety knowledge is left unstructured and unexploited. Knowledge and understanding of information within the H&S department are required not only to support H&S improvement initiatives, but also to aid in the definition of an integrated risk control system in order to maintain cohesiveness. The proposed conceptual framework introduces the idea of a digital-twin driven concept and can help decision-makers assess complex construction situations by utilising knowledge of several key parameters from an automated up-to-date risk factor database.

Objective 3: Develop a H&S information flowchart, as well as an information management and control framework for the H&S department to incorporate in-depth data communication and cross-departmental connection.

This objective addresses the variations in H&S information management caused by misinformation and misunderstanding of the information flow. There is still a gap regarding the specific flow management performed by the H&S department as part of their management function. To overcome this limitation, The developed frameworks strategically present a thorough digital representation for information management and control, which is capable of assisting high-level H&S management.

The above-mentioned approaches and applications demonstrate a continuous improvement in hazard control, construction digitalization, and safety information management. Within these broader studies, the main objective of this study is to develop an occupational safety hazard analysis and control with the proposed integrated frameworks to assist the construction industry in identifying key safety attributes and high-level management health information through digitalization. Key safety attributes include causes of hazards, physical and environmental demands, sources of injuries, ergonomic characteristics, level of injuries and their corresponding risk controls and preventions. The information management and control would also aid in prioritising departmental information flow in the H&S division. The study is expected to document the information flow within the H&S department at the facilitated management level and propose a framework for departmental information management and control to provide and enrich efficient information control along with effective support for enterprise decision-making. If the proposed comprehensive system can be fully implemented, the enterprise would expect to produce better informed and effective construction risk reduction solutions.

1.3 Thesis Organization

This thesis is composed of 5 chapters in total.

CHAPTER 1: Introduction. This chapter presents the research background, motivations, and objectives of the proposed research.

CHAPTER 2: Literature Review. This chapter provides a thorough review based on existing literature on the construction industry along with research gaps findings, including an overview of safety-related documents and its digitalization, current studies on PDA, SOP, and injury reports, attribute-based and risk identification approach to construction safety and H&S information flow, information management and control.

CHAPTER 3: Methodology. This chapter proposes the methodology for risk control and management framework, the foundation for the digitalization of risk assessment and control as well as digitalized H&S information management is also discussed in this chapter.

CHAPTER 4: Implementation Results, Discussions and Case Study. This chapter presents and discusses the comprehensive blended analysis results as well as three novel digital based frameworks.

CHAPTER 5: Conclusions. This chapter provides a summary of the conclusions that are drawn from this research. The limitations and future research directions are also outlined in this chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview of Safety-related Documents and its Digitalization

Occupational hazards assessment and risk identification in construction industry have become primary tasks as part of the prevention schemes and key safety documents are often recorded in the investigation reporting system to improve safety performance and regulate safety actions. Safety related documents generally include health forms, timesheets, requests, job procedures, assessments, injury reports and facilitating extraction of applicable safety requirements from these documents has merged as a significant subject in the construction safety domain (Wang 2013). Based on this recognition, there have been considerable research studies exploring the safety related documents in reporting system to ascertain the underlying risk factors (Abdelhamid et al. 2000; Chi and Han 2013; Li et al. 2019; Nesmith et al., 2013). Additionally, industries have attempted to adapt digitalization of safety documents to effectively integrate risk evaluations. Teizer et al. (2007) demonstrated that construction workers could be provided with layers of protection by implementing emerging risk controls and related documentation in H&S through a digital platform. An incident database system addressed by the safety management system (SMS) is proposed by Basso et al. (2004) to record factors causing incidents based on safety documents. A similar type of digitalization to safety information management system based on risk analysis is also proposed for construction enterprises (Yi 2019). According to Zweber's approach (Zweber et al. 2017), the digital twin-driven architect is built upon a data model that incorporates all applicable functional details that characterize aspects of the system with specific tasks during the system life cycle. More recently, Paolo and his research members present a joint Double Helix model together with the systems engineering and information technology to capture the temporal dynamics of the environments by using the data in both offline simulations and the online operational phase (Paolo et al. 2020). Example of another manufacturing application includes a proposed unified digital twin-driven framework for the real-time monitoring as presented in Yassine's work (Yassine et al. 2019). The overall digital twin-driven platform is segmented into individual digital twin-driven to satisfy factory-wide objectives by considering elements such as physical topology, system processes and machine processes. Among other approaches, Zhang et al. (2019) also presents a five-dimensional fusion model of a digital twin-driven virtual entity that specifically targets the robotics-based manufacturing system by reconfiguring production tasks. The combination of big data, digital technology, and construction has improved the H&S of construction workers. Integrating safety-related documents into occupational safety hazards and risk management with digitalization provides a decision tool for determining the appropriate incentive to invest for injury prevention.

2.2 Current Studies on PDA, SOP, and Injury Reports

The current study on physical demand analysis (PDA), as one of the safety documents, tends to provide a systematic approach to station-by-station injury data collection through an evaluation of worker's physical body posture demand, sensory, strength, job overview and environmental demands (Li et al. 2019, Li et al. 2015). An example of another PDA based application includes a proposed physical demand information form for systematic methodology for quantifying and evaluating all physical and environmental demand components of a job's essential and non-essential tasks as presented in Gagne's work (Gagne et al. 2022). The standard operating procedure (SOP) also serves as a critical safety document for maintaining consistency in the degree of work performance (Khairunnisa et al. 2020). By specifying necessary procedures on various types of work environments and workspaces, this will secure construction workers from occupational hazard events. Proper monitoring of SOP compliance also helps to reduce inadequate behaviours (Verma et al. 2014). Additionally, in one of the earliest accident report studies, Brazier (1993) introduced the benefits of using injury reports for the prevention of accidents after investigating numerous industry reports. Another study from Chi et al. (2013) incorporated Occupational Safety and Health Administration (OSHA) injury reports with Heinrich's domino theory to explore the accident causations. By considering the integration of both digitalization and smart usage of these safety related documents, this would bring significant benefits to the automatic risk assessments and ultimately secure construction safety.

2.3 Attribute-based and Risk Identification Approach to Construction Safety

A few approaches can be used to identify risks by using safety related documents. The existing literature on the construction industry with attribute-based approach reveals comprehensive guideline and knowledge baseline of safety risk quantification. The attribute-based risk model for measuring safety risk is first introduced in 2012, the study scrupulously utilized an attribute-based identification with identified 34 fundamental attributes over 300 injury reports from the national database (Esmaeili and Hallowell 2012). The model provided detailed relative risk values based on frequency and severity with content analysis (Baradan and Usmen 2006). However, the safety risk quantification strategy of their framework was limited by the unit quantified link where all subsequent result and analysis are based on the unit risk that only focuses on the frequency and severity of the incident. To address this limitation, this study utilized degree of risk assessment to further quantify risks associated with frequency of exposure, incident probability and potential consequence. Another quantifying attribute-based risk model has also been proposed by Villanova (2014) to conduct risk evaluation for industrial construction projects. The method takes into account the list of alternative upstream attributes with lower risk for safety risk analysis which allows assessment of complex situation. Yildiz and his team (2014) further applied the risk mapping tool based on defined risk attributes to facilitate decision making on risk rating.

More recently, numerous methods have been proposed to assist in the identification and grouping of keywords for risk prevention and safety prediction strategies. Risk perception technique is adopted by Hallowell and his team (2020) to forecast the effect of risk identification based on the integration of leading indicators, precursor analysis, safety climate and risk assessment. The leading indicators presented in Hallowell's research help to contextualize the findings on the quantity of safety management and activities (Hallowell et al. 2020). Suraji and his team (2001) emphasized the complex interaction of key risk causation grouping based on an empirical incident model. Similarly, Mitropoulos and Namboodiri (2011) presented an observational method for measuring safety risks based on observable risk factors and various causations. Other studies have attempted to classify applications of various key factors by using distinctive visual representation of all significant factors influencing injury analysis (Kartam et al. 2000; Haslam et al. 2005; Aksorn and Hadikusumo 2008; Li et al. 2013; Desvignes et al. 2014). Many studies have also investigated the quantitative computing techniques of grouping keywords. Ciarapica (2009) and his team have considered the probability and consequences of injuries based on soft computing techniques for identifying general factors. In comparison with commonly used correlation models, the study shows strategies for handling the inter-relationships among different variables. However, given that most models of key words grouping involves various combinations of causative risk factors, data mining techniques are likely to retrieve massive amount of data and uncover even more hidden patterns in the field of safety quantification analysis. In addition to this concept, both Liao and Perng et al. (2008) and Tam et al. (2004) concluded important characteristics of occupational hazards examination by using data mining analysis on key factors contributing to construction related injuries in Taiwan. Another existing method based on automated content analysis is also useful for providing the overall data mining methodology that focuses on the relationship between text mining and safety assessment, and specifically how keyword detection can be used for contributing to the fundamental hazardous attributes (Tixier et al. 2016). Attribute-based and risk identification approach allows for the unique definition of the baseline characteristics of construction risk attributes. Given that existing models only discuss quantified interactions between various exposure frequencies and their relationship to the degree of probable consequence, this thesis focuses on a concise risk quantification based on exposure frequency, incident probability, and potential consequence for risk assessment.

2.4 H&S Information Flow, Information Management and Control

OHS experts are compelled to concentrate on the strategic management of their organizations as construction sector becomes more globally concerned in terms of its health protection and worksite safety. To reduce OHS injuries, illnesses, and fatalities, greater emphasis is being placed on strategic management and "beyond compliance" strategies (Vladimirovna et al. 2014). An integrated information management for construction safety and health information flow could detect and reduce occupational risks in which may promote continuous improvement, rationalization, and dependability of projects, processes, and services. New methods of H&S integration with information flow, information and control management are therefore crucial in this setting. Numerous studies have identified the use of digital information management system as an effective technology for performance improvement and H&S data measurement and control. Several aspects of construction H&S have been improved using online systems, including safety training and education, risk identification, safety monitoring and evaluation, and safety inspections (Dodge Data and Analytics 2017).

Shimada et al. (2014) comprehensively introduced an integrated approach for occupational safety and health based on the business process model of engineering activities. The team pointed out that widely used and generalized structures of activities and information flow related to construction activities should be developed throughout the stages of company performance management. Redinger et al. (2011) also emphasized the importance of deploying the occupational H&S management systems as a powerful risk management tool. To play a key role in their organizations, OHS professionals must understand existing management system information flow and approaches, how to develop and implement such information systems, and the conformity-assessment structures that affect these systems (Redinger et al. 2011). Moreover, Zhou and his team (2012) proposed an online system that employs artificial intelligence to assist H&S information capture and analysis, as well as support decision making through risk identification and assessment. Through a prototype developed for the pattern execution and critical analysis of construction site-space organization, Zhou et al. (2012) analyzed both topological and geographical information flow which includes product and process geometrical data, activities schedule information as well as execution patterns library and safety work rate data. Similarly, Cheung et al. (2004) designed a construction safety and health monitoring (CSHM) system as an analyzer of potential risks and hazards, as well as a warning sign for construction activities that require immediate corrective action. The web based CSHM allows for remote access of information, quick data collection, retrieval, and documentation by leveraging benefits of digitalization. The design also included a knowledge base to allow for expert advice and instructions of company's H&S recommendation. Another web tool prototype is presented to aid in the evaluation of information flow between potential designers, contractors, and coordinators within the construction project (Yu et al. 2009). To support safety decision-making mechanisms, a knowledge-based decision model based on rule-based reasoning, case-based reasoning, and hypertext technology is proposed to facilitate information flow and control. Yu and his research's work (2009) aimed at various levels of project, product, process, and operation safety management in the construction industry. Online databases are used to assess the competence of various stakeholders and management control, with research tools designed to enable project safety information queries and communication within major corporations.

Management Information System (MIS) is a more generic system comprised of humans, computers, and other components that can collect, transfer, store, protect, and use information (Zapalac et al. 1994). In recent years, an enterprise management information system (EMIS) is also established to determine enterprise development strategy, reengineer business processes, analyse business data, and define system functions for use in the enterprise information flow (Kai et al. 2004). Among the challenges by the construction division in the industry, suitable information management system for construction engineering is adopted as a productive tool for

H&S integration. To accurately reflect the construction dynamics in the data information management, integrated and control management information system of construction engineering project enterprises groups is constructed for data planning and system framework design of entire enterprise (Qi et al. 2009). Additionally, Yin (2019) proposed the organic integration of information management and civil engineering construction supervision, thereby effectively ensuring civil engineering construction management safety. According to the information management system, historical and current data from each department will be centralised for leaders to make decisions (Bo et al. 2002).

Review of literature on the above integrated H&S information management showed that, although there is a lot of scholarly literature on construction information flow and control management of the entire enterprise, there is still a gap on the subject of the specific flow management performed by H&S department as part of their management function. Human factors, such as potential misinformation flow and linkage, misunderstanding when interacting with other departments, and unrealized human mistakes when recoding data, were not considered in computing strategic data control within the H&S department. Knowledge and understanding of information within H&S department are necessary not only to support H&S improvement initiatives but also aid the definition of an integrated risk control system for the purpose of maintaining consistency. The solution to this limitation consists in providing detailed information flow that has specific focus on the H&S department, including information mapping, hierarchical integration of the safety data management at each level of the department to realize the consistent and collaborative OSH management. To overcome this limitation, this study also proposed a H&S department information flowchart as well as a management and control framework that specifically address for H&S

department. By combining the above concepts and frameworks, the overall integrated construction information management can ultimately promote overall economic, safety and social sustainable development.

CHAPTER 3: METHODOLOGY

The detailed research process flow that used in this study is summarized in Figure 1. To present a blended analysis of the occupational safety hazard and risk assessment with digital information to assist enterprise H&S information digitalization, construction-related data is examined in this study as shown in the inputs section. Detailed inputs include: (1) In-plant/Field job demand and incident report; (2) Physical demands analysis (PDA); (3) Standard operating procedures (SOP); and (4) Categorized risk factors assignments.

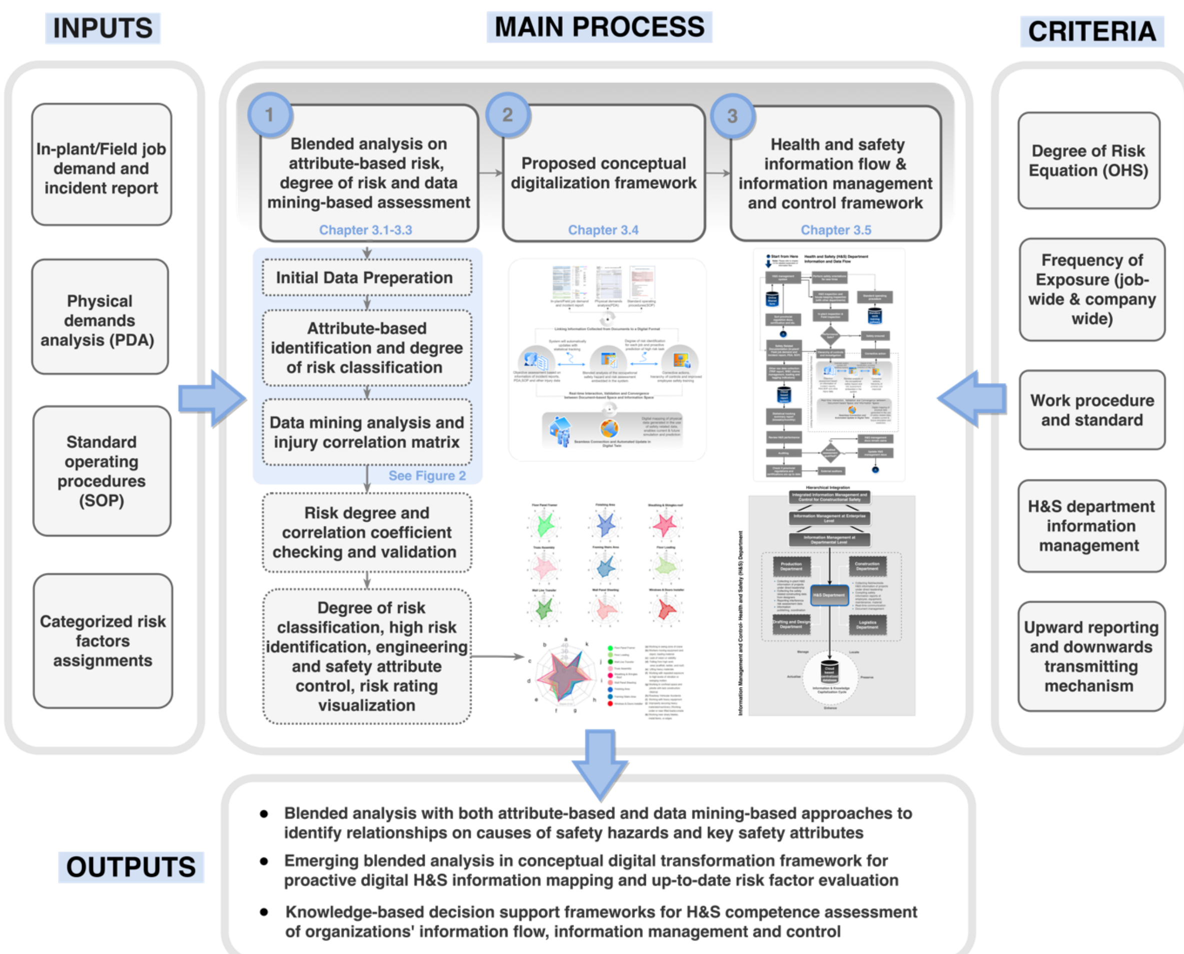


Figure 1: Detailed Research Process Flow

As illustrated in Figure 1, the main proposed research process flow can be divided in to three sections as follows:

1. **Blended analysis on attribute-based risk, degree of risk and data mining-based assessment:** The development of this blended analysis contributes to the achievement of the research objective 1. The main tasks include both attribute-based and data mining approaches that are being developed to assist the construction industry in identifying causative factors of hazards and risky safety attributes. The methodology development is described in greater detail in chapter 3.1, 3.2, 3.3 and can be illustrated as shown in Figure 2. The performance of the risk degree and correlation coefficient is also validated and cross checked with the previous studies and literatures. Detailed implementation, classification and discussion are provided in chapter 4.1, 4.2 and 4.3.

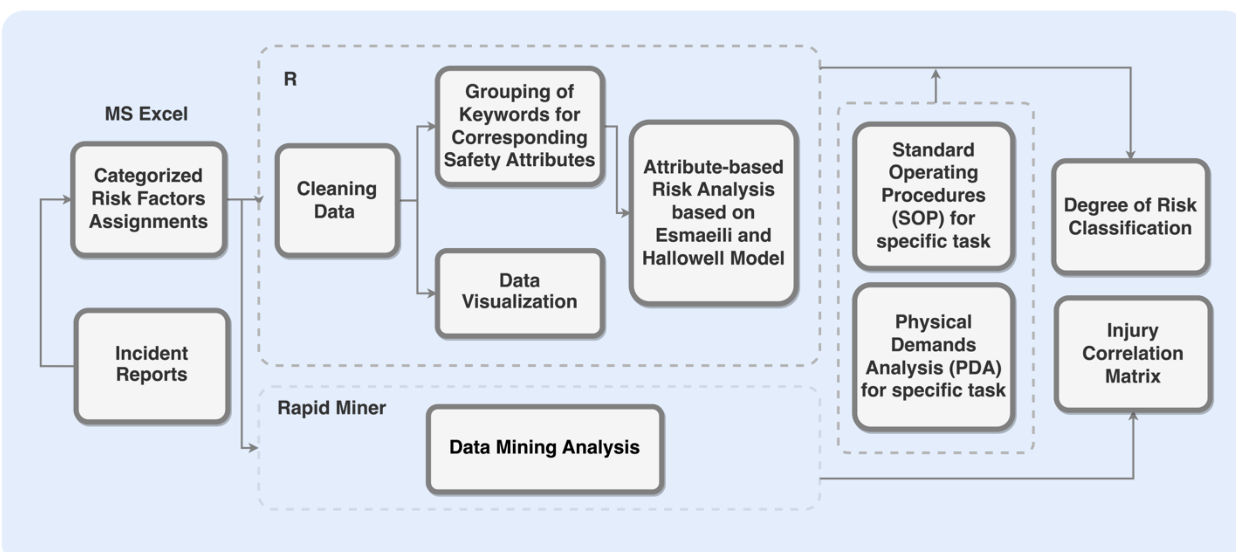


Figure 2: Detailed Research Process Flow on chapter 3.1 & 3.2 & 3.3

2. **Proposed conceptual digitalization framework:** The development of this method serves the research objective 2, where the method proposes a conceptual H&S digitalization framework that integrate both blended analysis of the occupational safety hazards and risk assessment and digital twin-driven design. The detail of the proposed methodology is shown in Figure 4 and a comprehensive description is provided in chapter 4.4.
3. **Health and safety information flow & information management and control framework:** The development of the information flow and management method is proposed to fulfill the research objective 3. One of which refer to the information flowchart that specifically focus on H&S division while the other refer to the high-level H&S information management and control. The proposed method will be further detailed in an application to collaborative industry partner. The reliability and effectiveness of the method is also validated through case studies with the industry partner and detailed implementation can be further explored in chapter 4.5.

In addition, as can be seen from Figure 1, the criteria module is also identified for the proposed risk analysis and frameworks, which include 5 parts: (1) Degree of Risk Equation (OHS); (2) Frequency of Exposure (job-wide & company wide); (3) Work procedure and standard; (4) H&S department information management; and (5) Upward reporting and downwards transmitting mechanism. Detailed description and implementation are revealed in the following methodology.

3.1 Initial Data Preparation and Process Flow

H&S control is a multidisciplinary field concerned with guarding the safety and health of people who are engaged in work or employment. The blended analysis approach developed in this study first provides a quantitative evaluation on the occupational risk (See section 1 Figure 1), which is used to assess the level of injury and the occurrence of serious injury cases. The detailed research process flow that used in this section is summarized in Figure 2 as previously shown. This provides the foundation for digitalization of risk assessment and control as well as digitalized H&S information management

Objectives of this section are fulfilled with the attribute-based risk analysis and the data mining analysis, emphasizing the dominant cause of injuries and its interrelation with the human body. Incident injury cases are first collected from construction-related injury reports. To identify the related cause-and-effect considerations and keywords, the study then characterizes factors based on each performance of the injury reports in the following categories: motion injuries, equipment and source of injuries, area of injuries, safety hazards and ergonomic risk factors with a sample report summarized in Table 1. This thesis first used R programming language for the quantitative risk evaluation. The R programming language (R Core Team 2014) allows classification, grouping of keywords, visualization of dataset and specifically involves comprehending and analysing human-produced texts. This study employed the R programming language for preprocessing database as mentioned in the Figure 2, specific cleaning process included: upper-case and lower-case conversion as R is case sensitive, omission of N/A variable and extra space elimination. In addition, to effectively perform grouping of keywords for corresponding safety attribute within certain level of timeframe as depicted in the Figure 2, R programming language was also used to achieve this goal. Detailed attribute categorization used for the study (based on the pre-assigned categorized factors) and R coding information are presented in Appendix A.

In the pre-processing stage, these factors are assigned to each reported case to ensure an in-depth understanding of the injury outcome and can be shown in the categorized risk factors assignments block in Figure 2 (Section 3.1). The availability, quality and reliability of the data will also be assessed (Fayyad 1996). After completing the initial data management, each corresponding assigned risk factor is prepared for the quantification of the risk based on the attribute-based framework of Esmaeili and Hallowell (2012). The R programming language is mainly used for classification, grouping of keywords and visualization in this section.

Table 1: Pre-assigned categorized factor for cause-and-effect relationship and sample injury report

Motion Injuries	**Equipment and Source of Injuries**	**Area of Injuries**	**Safety Hazards**	**Ergonomic Risk Factors***
Carrying, craning, driving, exiting, falling, kneeling, lifting, nailing, pulling, pushing, slipping, swinging, walking	Crane, falling object, foreign object, hammer, heavy object, ice, ladder, machine, metal item, mud, nail gun, propane, trailer, vehicle, wrench	Ankle, arm, back, chest, face, finger, foot, hand, head, hip, knee, leg, shoulder, wrist	Inadequate maintenance, Inadequate clearances, Inadequate Guards and protection, Inadequate PPE	F/P, Forceful Exertions, poor posture, R/P, repetition, vibration

Report	**Motion Injuries**	**Equipment and Source of Injuries**	**Area of Injuries**	**Safety Hazards**	**Ergonomic Risk Factors**
Was lifting a lift point, from my jig to install on stair cage and felt something tweak/pop just above my tail bone in the middle of my spine	Lifting	Heavy Objects	Back	N/A	F/P

**Noticed that F/P in the following table can be interpreted as ergonomic risk factors that contain both forceful exertions and poor posture and R/P can be viewed as ergonomic risk factors that represent both repetition work and poor posture.*

3.2 Attribute-based Risks and Degree of Risk Assessment

To determine the magnitude of safety, attribute-based risk analysis model together with the degree of risk assessment is developed (See Figure 2) and in total 11 safety attributes were identified with modifications. Three major groups including Motion and Physical characteristics, Workstation and Jobsite and Equipment, Material and Source have been classified in regards with all corresponding safety attributes. Next, R performs the grouping of keywords for corresponding safety attributes for the realization of each attribute. For example, in attribute "Workers moving equipment and object, loading material", key motion injury words such as "carrying", 'pulling" and "pushing" are utilized and contributed for this corresponding attribute. After identifying the proper list of safety attributes, a method employing quantification of risk value is then applied. Instead of considering the quantification method introduced by Baradan and Usmen (2006), this study adopts the evaluation based on the degree of risk equation presented by the Government of Alberta OHS Program, illustrated by Equation 1 (OHS 2011).

[1] Degree of Risk = Frequency of Exposure (FE) × Incident Probability (IP) × Potential Consequence (PC)

Equation 1 uses three factors to analyze the risk: Frequency of Exposure (FE), Incident Probability (IP) and Potential Consequence (PC) of loss. As shown below in Table 2, four different levels of FE per company are pre-defined based on the safety standards adopted from OHS (2011) and is used in the ultimate degree of risk calculation. However, FE is adjusted to have a more objective and automatic assessment (See chapter 3.2.1). Based on the following table, in total 4 levels of classification are defined. Frequency of Exposure (FE) describes the relationship between the level of exposure (daily, weekly, monthly or occasionally) and individual competence on task. Incident Probability (IP) represents how often (probable, occasional, remote or improbable) an incident is occurred. Potential Consequence (PC) of loss demonstrates detailed injury type (severe, substantial, minor or occasionally) due to the incident. Thorough explanations are presented in the following sub chapters.

Table 2: Detailed level of classification for FE (per entire company), IP and PC (OHS 2011)

Detailed level of classification	**Frequency of Exposure (FE)**		**Incident Probability (IP)**		**Potential Consequence (PC)**	
	Frequency	**Description**	**How often**	**Description**	**Injury Type**	**Description**
4	Daily	Task is performed one or more times a day	Probable	May happen at least once a month	Severe	Lost Time Claim
3	Weekly	Task is performed once a week	Occasional	May happen once every 1-4 months	Substantial	Medical Aid
2	Monthly	Task is performed once a month	Remote	Not likely to happen, but possible once every 4-12 months	Minor	First Aid Injury or Property damage
1	Occasionally	Task is performed less than once a month	Improbable	Not likely to happen	Occasionally	Near Miss

3.2.1 Frequency of Exposure

In Equation 1, FE represents the level of exposure to the hazard at the workplace when individual completes activities by considering with each safety attribute. In most cases, this FE can be treated as the frequency in which an employee performs the task. Noticed that this factor (as per entire company) is more abstract in its nature because activities and hazards can sometimes be both complex and dynamic. Furthermore, FE varies from one job to another when considering its relationship with each defined safety attribute. For example, a crane operator mainly focusses on craning activities during the shift, however, operator may also encounter other sub-tasks and not only controlling the crane and working under the swing zone but also loading and securing material as a second characteristic. To this end, human judgment based on safety standard is first adopted to assess the FE per entire company in the literature as illustrated in Table 2. Next, FE that specifically addressed for each task (per job) is then selected to reflect further risk level identification by employing the knowledge of several key parameters from PDA and SOP. This adjustment allows reliable prioritizing of identified risks (per each job) through objective assessments.

A case result on adjusted frequency of exposure for floor loading task is illustrated in Figure 3. The standard operating procedure is essential as a consistency measurement in the level of work performance. In this case, SOP help defines whether floor loading task or operation encounter with defined safety attributes characteristics. Next, PDA provides concise motion data in terms of activity force, frequency of workday as well as other job critical demands. For example, based on key contribution indicated by PDA and SOP, cranes (Overhead and Mobile) and securing straps are mainly used to hoist prefab walls and flooring panels and therefore perceived a score of 4 for safety attributes on swing zone area and heavy material securement. Manual handling tasks from PDA provided additional information on frequency of workday, the frequencies on PDA are ranked from most frequent to lease frequent as “constant”, “frequency”, “occasional” and “rare”. A constant 67-100% of the shift associated with forceful gripping, grasping, and pinching is recorded and therefore was assigned with a score of 4 with moving equipment and object. Additionally, there is no explicit consideration regarding high level vibration hand tool, sharp blades items and roadway vehicular condition from floor loading, thus suggesting a score of 1. Other key useful information including vison requirements, high heights condition and basic ergonomic risks are also used to determine adjusted FE score. As a result, standard operating procedure together with physical demand analysis presents a comprehensive overview highlighting both physical frequency data and environmental assessment for each workstation, therefore provides an objective assessment of frequency of exposure with each job.

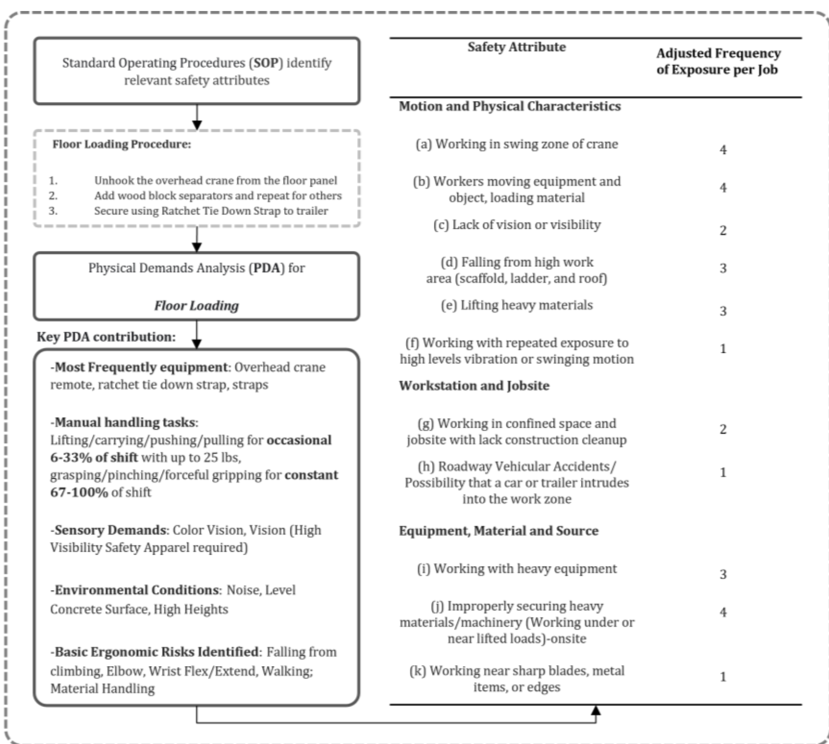


Figure 3: Sample case result of frequency exposure determination for floor loading task

3.2.2 Incident Probability

In Eq. 1, IP in this study indicates how likely the exposure (or identified safety attribute) will result in loss. The loss in this case can be viewed as any injury, illness, property damage, poor work quality or even lost production. As illustrated in Table 2, the scale of the IP is modified within the appropriate time frame with respect to the given database. In addition to the scaling modification as mentioned before, the IP factor also considers the identified safety attribute without current occupational control mechanisms to perform a consistency check for each listed attribute.

3.2.3 Potential Consequences of Loss

Similarly, PC in this study represents the severity of the loss for each attribute at the workplace if the exposure is not controlled. As provided by Table 2, PC can be easily quantified by counting number of injury types (Lost Time Claim, Medical Aid, First Aid, Property Damage and Near Miss) that correspond for each safety attribute. To eliminate inconsistencies among types of injury identifications, if same number of PC levels has been detected, then the one that contained more severe case is chosen for a consistent purpose.

Finally, by conducting all levels of classification for each safety attribute based on Eq. 1, the degree of risk can then be determined by multiplying the scores of the three factors together according to the previous formula (Eq. 1). Noticed that to ensure the consistency of the judgments in all degrees of risk, both company-wide and job-wide frequency of exposure determination can be defined within the four different levels. In accordance with measurements provided by the OHS, each degree of risk can be classified into certain range of levels or scores. Low risk (L) refers to the degree of risk from score 1 to 9, this suggests that continued operation is permissible with minimal controls and recommends track the risk and take steps if the risk level increases. Score of 12 to 27 represents medium risk (M) and it often take timely action to implement appropriate controls to lower or minimize it. Score of 32 to 64 indicates high risk (H) and it requires immediate action. This classification will help to establish the priority for the implementation of control measures, and more specifically, to identify relative safety attributes with medium or high risks.

3.3 Data Mining-based Assessment

Data mining analysis is then conducted following the attribute-based risk assessment model to further reveal the associations, visualization, and correlation coefficient of potentially risky activities and construction hazards based on pre-assigned categorized factor (See Figure 2) for each injury cases. This thesis used RapidMiner for further exploration on the association between each safety keywords. RapidMiner is preferred due to its high-quality module features, and it is applicable in many contexts of text analytics. The focus of this section of the thesis is mainly on text-mining and text-association analysis by using RapidMiner with matrix visualization. In this phase, this research utilized different modules and operators to identify and reveal further correlation among associated risk factors. As illustrated in Figure 2, after assigning categorized risk factor for each reported case, dataset is transformed into the RapidMiner to relate all these attributes together for each case. The schema provides the integrity of the data and eliminates redundancy. In addition, it links all pieces of information together which results in finding high risk industry and occupation conditions to be able to provide administrative and engineering control actions to mitigate their risks. Moreover, it stores data in convenient metadata to be ready for further processing and information extractions. Detailed data processing includes (1) handling text columns, (2) preparing data for correlation calculations, (3) one-hot encoding performance (4) removing ineffective columns, i.e., columns with constants, (5) sampling data down based on the number of attributes, (6) ordering columns alphabetically and finally (7) creating the actual correlation matrix. The correlation matrix operator in RapidMiner can be used to calculate the correlation between all provided variables. The output weights are normalized, so that the highest score is 1 and the lowest is 0.

3.4 Digitalization Adoption

This thesis also proposes a conceptual model for integrating safety related documents and blended analysis of the occupational safety hazards and risk assessment into a digital format as indicated in Figure 1 section 2. The proposed framework (Figure 4) is intended to aid decision makers during construction evaluation of the impacts for prioritized risk factors on a digital platform, which can be further utilized in a digital twin-driven design that enables future risk prediction and digital mapping of physical safety related data. The digital twin-driven manufacturing tends to fulfill any requirements throughout a sustainable intelligent manufacturing and construction industry. However, the concept of digital twin-driven implementation is broad and more complex in its nature. Each stage process needs to be conducted in a standardized manner and requires an integration that incorporates all aspects from a company. The current proposed framework is intended to draw attention to the importance of H&S digitalization. Detailed discussion and implementation are provided in chapter 4.4.

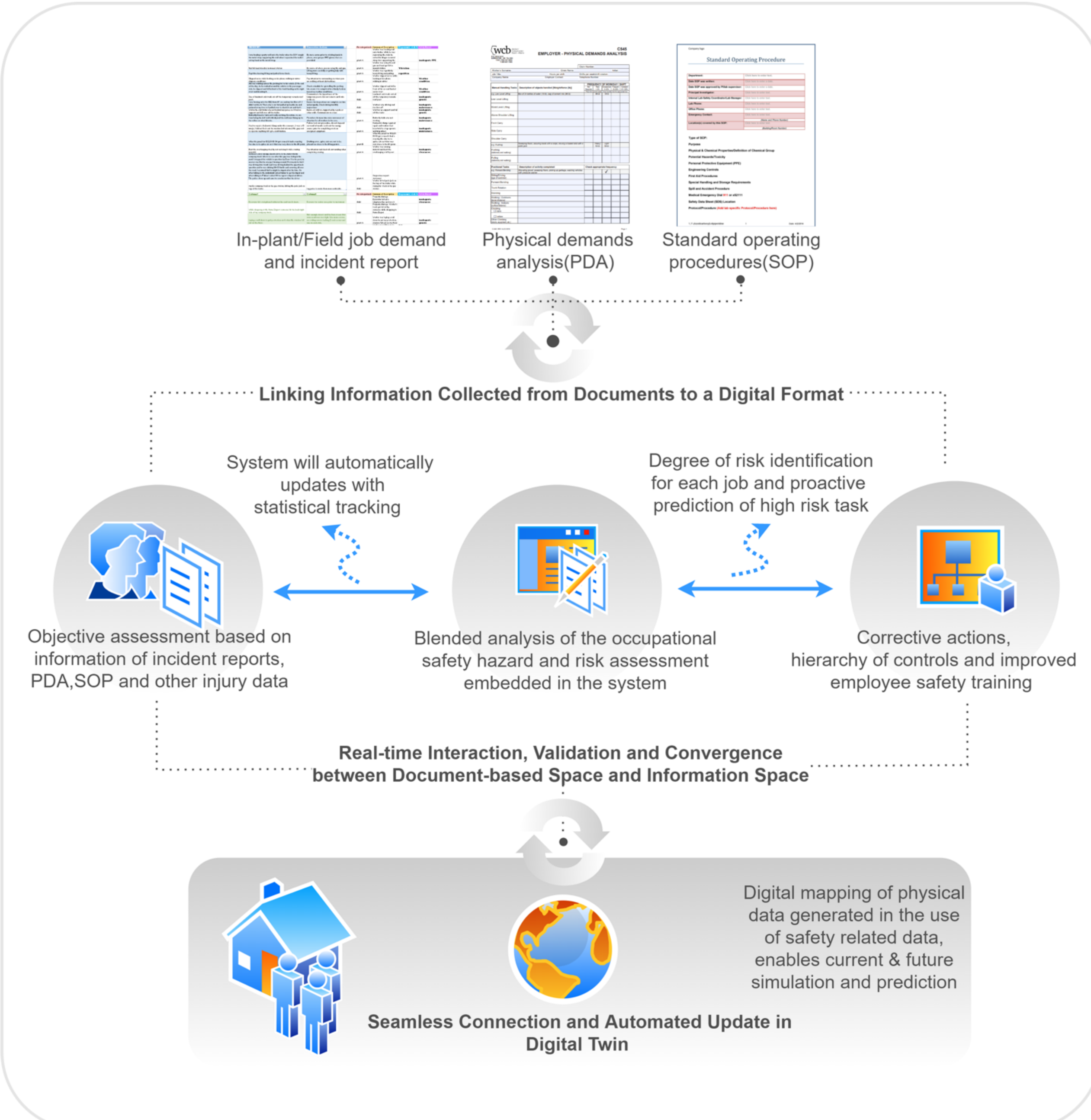


Figure 4: Framework of proposed digitalization future model

3.5 Digitalization on H&S Information Flow, Information Management and Control

The traditional strategic management in the civil engineering construction process is mostly manual supervision, which inevitably affects the quality of safety and health information management due to a lack of human resources and the lack of professional skills of experts (Yin 2019). A H&S department information flowchart, as well as information management and control framework for H&S department are also proposed as indicated in Figure 1 section 3 to support safety decision-making mechanisms and achieve consistent and collaborative OSH management. Information mapping and hierarchical integration of safety data management at each level of the enterprise are established.

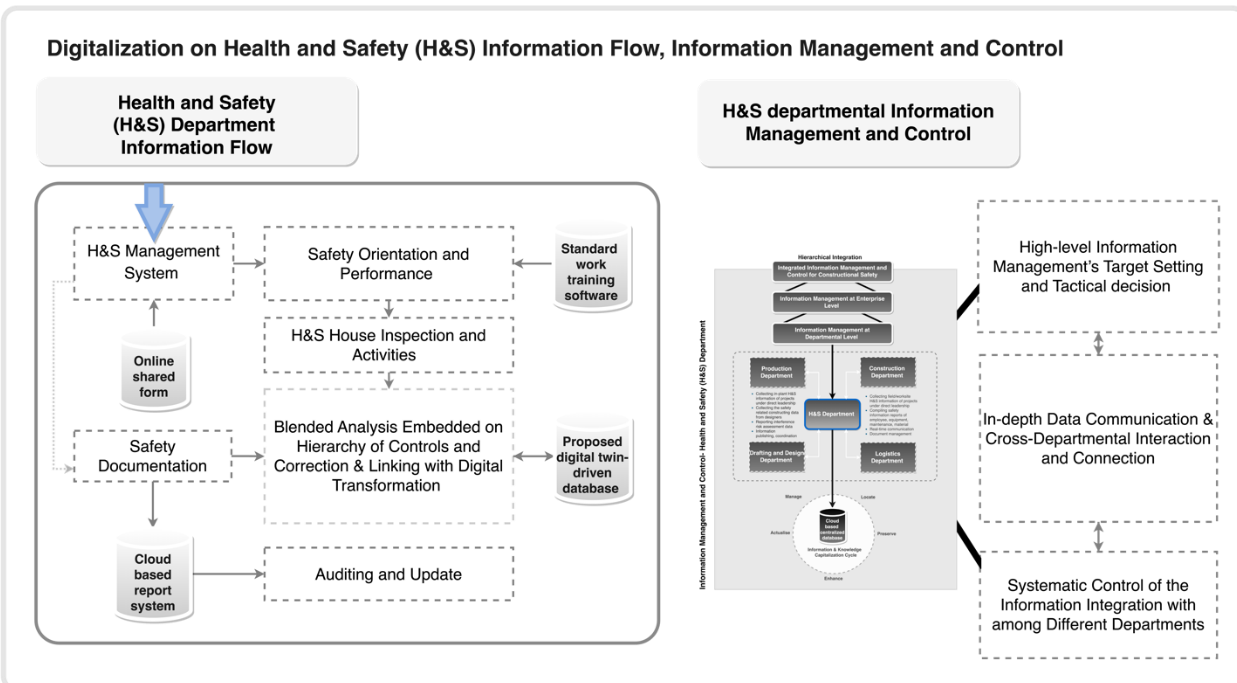


Figure 5: Proposed methodology on digitalization on H&S information flow, information management and control

The current proposed methodology is demonstrated below in Figure 5, a general H&S department information flow on the left will be further elaborated in chapter 4. For the efficiency of H&S information collection, retrieval and investigation, this information flowchart is applied because of its performance, scalability, and reliability. The main process and activities include (1) H&S

Management System; (2) Safety Documentation; (3) Safety Orientation and Performance; (4) H&S House Inspection and Activities; (5) Blended Analysis Embedded on Hierarchy of Controls and Correction & Linking with digitalization; and (6) Auditing and Update. Additionally, H&S departmental information management and control framework is also proposed to fulfill high-level information management's tactics decision for cross-departmental interaction and connection. The current proposed methodology is embedded in information management to capture and evaluate the outcome of H&S performance while also transferring tacit knowledge into explicit knowledge. Comprehensive discussion and implementation are provided in chapter 4.5 based on a case study result. With the implementation of the above application and approach, the results present a continuously improved digitalization, emphasizing productivity and efficiency and its interaction with the information flow.

CHAPTER 4: Implementation, Results, and Discussions

The proposed attribute-based risk and degree of risk analysis, the digital twin-driven framework and the H&S information management and control framework have been implemented in a local industry partner. To illustrate the analysis of the occupational safety hazards and risk assessment, the developed blended scheme of both attribute-based and data mining-based approaches were used to analyze 180 construction-related injury reports. To identify the related cause-and-effect considerations and keywords, the study then characterized factors in the following categories: 6 ergonomic risk factors, 7 safety hazards, 14 area of injuries, 15 equipment and source of injuries and 13 motion injuries summarized in Table 1. In total, 9 construction related jobs are examined, namely floor loading, floor panel framer, finishing area, framing stairs area, sheathing and shingles of roof, roof truss assembly, wall line transfer operator, wall panel sheeting and windows & doors installer. To properly utilize the analyzed information and results, a conceptual digitalization framework has been first implemented that integrates quantitative risk analysis to facilitate the safety risk assessment through the digital platform. Both H&S information management and control framework have been implemented with a local industry partner as a case study. To effectively reveal the overview of the H&S information flow, a complete description of H&S processes, activities and organization functions have been demonstrated using images and flow diagrams. To further explore the H&S information management and control framework, both upward reporting mechanisms and downward data transmitting mechanisms have been employed to characterise cross-departmental interaction and information flow through departmental management. The main results are stated as follows:

4.1 Attribute-based Risks and Degree of Risk Analysis (Company-wide)

Table 3 reveals all relevant degree of risk with detailed quantitative results in regard to the predefined safety attributes. Key administrative and engineering control from overall company perspective is also provided as shown along with each attribute. A list of 11 attributes were identified based on the attribute-based model of Esmaeili and Hallowell and three major groups (Motion and Physical characteristics, Workstation and Jobsite and Equipment, Material and Source) have been classified in regards with all corresponding safety attributes.

Table 3: Degree of Risk classification (company-wide) and controls for each safety attribute

Safety Attribute	Frequency of Exposure	Incident Probability	Potential Consequences	Degree of Risk Classification*	Administrative and Engineering Control
Motion and Physical Characteristics					
(a)Working in swing zone of crane	4	2	1	L	Operational policy
(b)Workers moving equipment and object, loading material	4	4	2	H	Avoid manual task and overexertion
(c)Lack of vision or visibility	1	2	2	L	Use of automatic detection
(d)Falling from high work area (scaffold, ladder, and roof)	3	3	3	M	Training, operational policy
(e)Lifting heavy materials	4	3	2	M	Lifting strategy
(f)Working with repeated exposure to high levels of vibration or swinging motion	4	4	2	H	Reducing excessive motions

Workstation and Jobsite					
(g)Working in confined space and jobsite with lack construction cleanup	4	3	2	M	Regular construction cleanup
(h)Roadway Vehicular Accidents	3	4	2	M	Proper Training
Equipment, Material and Source					
(i)Working with heavy equipment	4	4	2	H	Limits controls
(j)Improperly securing heavy materials/machinery (Working under or near lifted loads)-onsite	4	4	1	M	Securement and appropriate protection
(k)Working near sharp blades, metal items, or edges	4	3	3	H	Proper PPE

**L= Low Risk, M= Medium Risk, H=High Risk*

An overview of the degree of risk analysis based on the attribute model is shown in Figure 6. This figure is intended to reveal the distribution of each safety attribute with respect to its own determined risk degree from overall company perspective, and thus determines the level of risk and the magnitude of safety. By referring to the high-risk zone (light grey shaded area on the top) of the Figure 6, one can see that "workers moving equipment, and object, loading material", "working with repeated exposure to high levels of vibration or swinging motion", "working with heavy equipment" and "working near sharp bladed, metal items or edges" have the highest degree of risk and can be interpreted as the most hazardous attributes for the given construction-related injury cases. These findings are in align with the studies by Thomson (1996) where majority equipment related accidents associate with lack of compliance following OSHA regulations and safety standards. By examining at attribute individually, “working near sharp bladed, metal items or edges” was the most prominently risky attribute. Reducing the time, a given worker is exposed to this attribute tends to reduce the potential risk of injury.

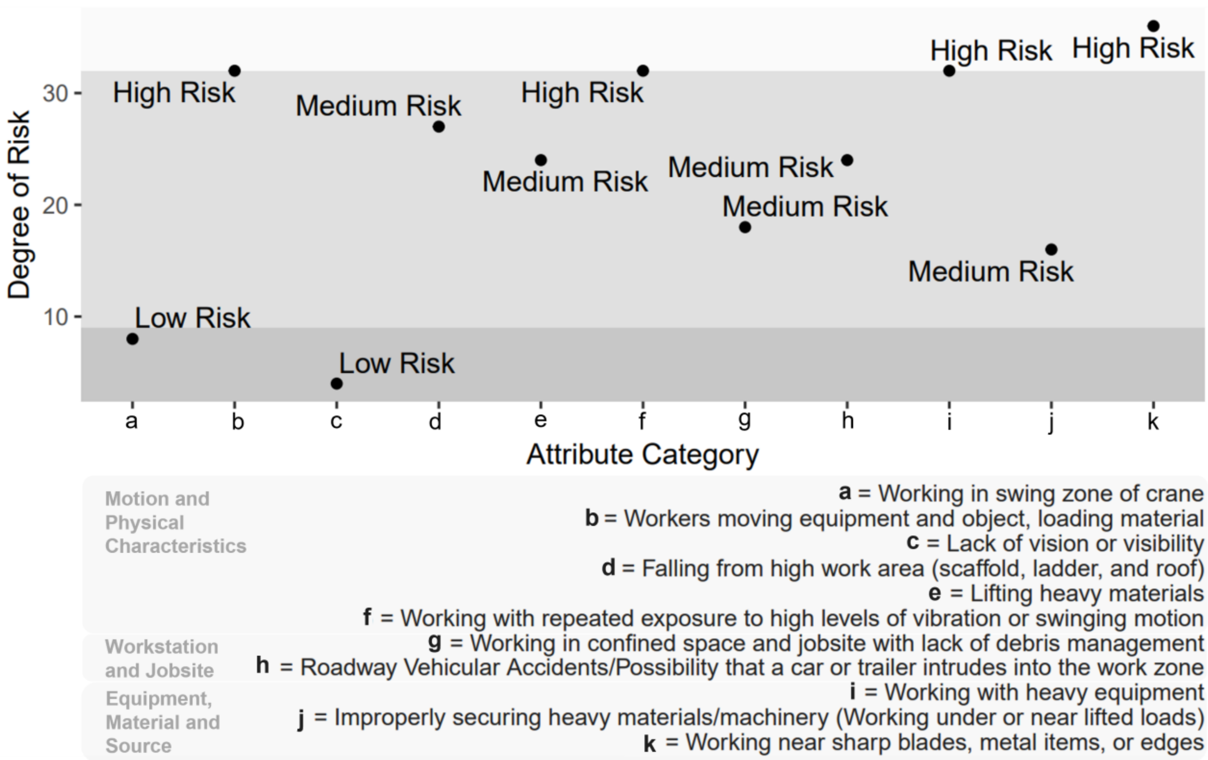


Figure 6: Detailed degree of risk classification (company-wide) for each attribute category

In parallel with this, visualization that implemented by R was also used to reveal the distribution of injury areas with respect to each categorized risk factor as shown in Figure 2, and thus provides further information of important baseline characteristics of the dataset. Figure 7 as following shows the relationship between area of injuries and four categorized incident risk factors. Distinct four categorized risk factors include previously defined motion injuries, safety hazard, ergonomic risk factors and equipment source of injury which specifies the elements of the most common source of risk that workers might encounter.

As can be seen above, “nailing”, “carrying”, “slipping”, and “lifting” motions resulted in most cases of motion injuries; “working from height”, “inadequate clearances” and “inadequate PPE” contributed to the major safety hazards; mostly frequently reported ergonomic risk factors were

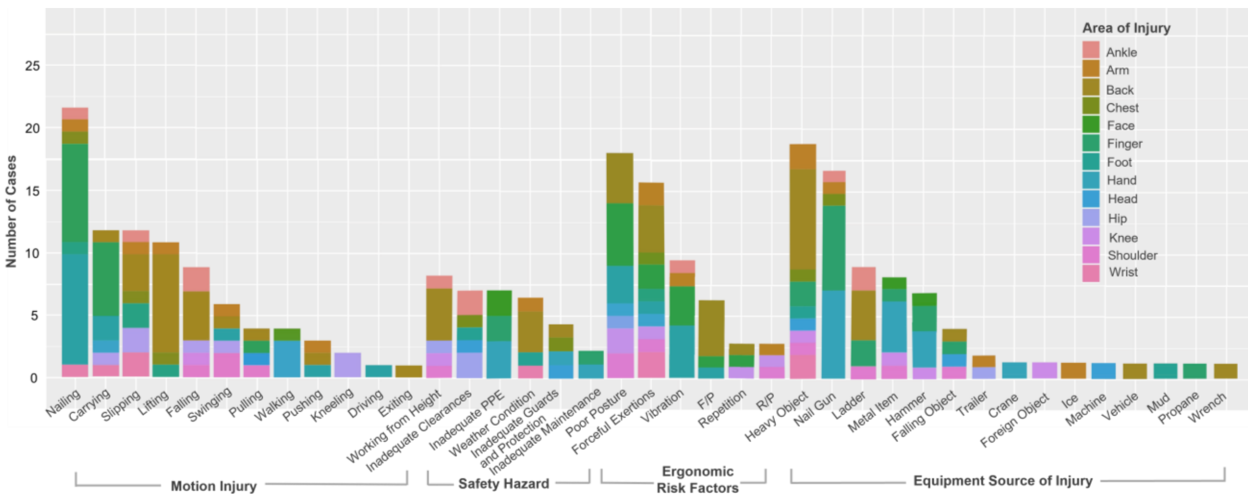


Figure 7: Number of cases in area of injury by ergonomic risk factors, equipment and source of injury, safety hazards and motion injury

submitted by “poor posture” and “forceful exertions”; major equipment source of injuries was resulted in “heavy object” and “nail gun”. By showing area of injury focused relationship, the visualization has also revealed an overlapping characteristic with the medium or high-risk safety attributes of Figure 6. These risks can be significantly reduced by practicing National Institute for Occupational Safety and Health (NIOSH) checklist. For example, Work-related musculoskeletal disorders (WMSDS) hazard identification checklist, risk factor report card, workstation checklist and task analysis checklist. Other suggestions that can reduce the body injuries include establishing control measures such as personal protective equipment (PPE): dust masks, proper gloves, hard hats, protective eyewear, and steel-toed safety boots (NIOSH et al. 2007)

4.2 Attribute-based Risks and Degree of Risk Analysis (Job-wide)

Construction projects are prone to diverse occupational safety hazards and risks with regards to different job tasks. By examining the internal factors that shape job-wide degree of risk, this additional approach explores the relevance between 9 construction related jobs and its risk perceptions associated with safety attributes. The following 9 construction related jobs are selected with regard to the typical job type distribution in the construction industry. This selection is established based on a common manufacturing construction enterprise that focus on panelized wall production and prefabricated home development. Typical jobs include Floor Loading, Floor Panel Framer, Finishing Area, Framing Stairs Area, Sheathing and Shingles, Roof Truss Assembly, Wall Line Transfer Operator, Wall Panel Sheeting and Windows & Doors Installer. Critical risk identification on each explored job can be easily displayed by radar charts as follow (See Figure 8). This visual identification was constructed by giving an axis of risk degree for each safety attribute variable and presents as a reliable method for comparing risk factors, defining priorities, allocating control recommendations, and evaluating performance. Depending on the job feature and environmental constraints, the decision makers may prioritize investing in specific job with larger risk magnitudes. Additionally, the data from multiple construction related job observation are plotted along axis, overlayed and connected to form a single summarized degree of risk radar chart for all different jobs (See Figure 8). By creating this single summarized radar chart, each job wide division's performance can be investigated and improved with human resource interaction, facilities renovation and training initiates. This supplementary summarized chart is made to prioritize relevant safety attributes placed for each job: "working near sharp bladed, metal items or edges" was perceived as the most significant overall risk category amongst all jobs followed by "falling from high work area", "workers with repeated exposure of high levels of vibration"," workers moving equipment, and object, loading material" and" working with heavy equipment", respectively.

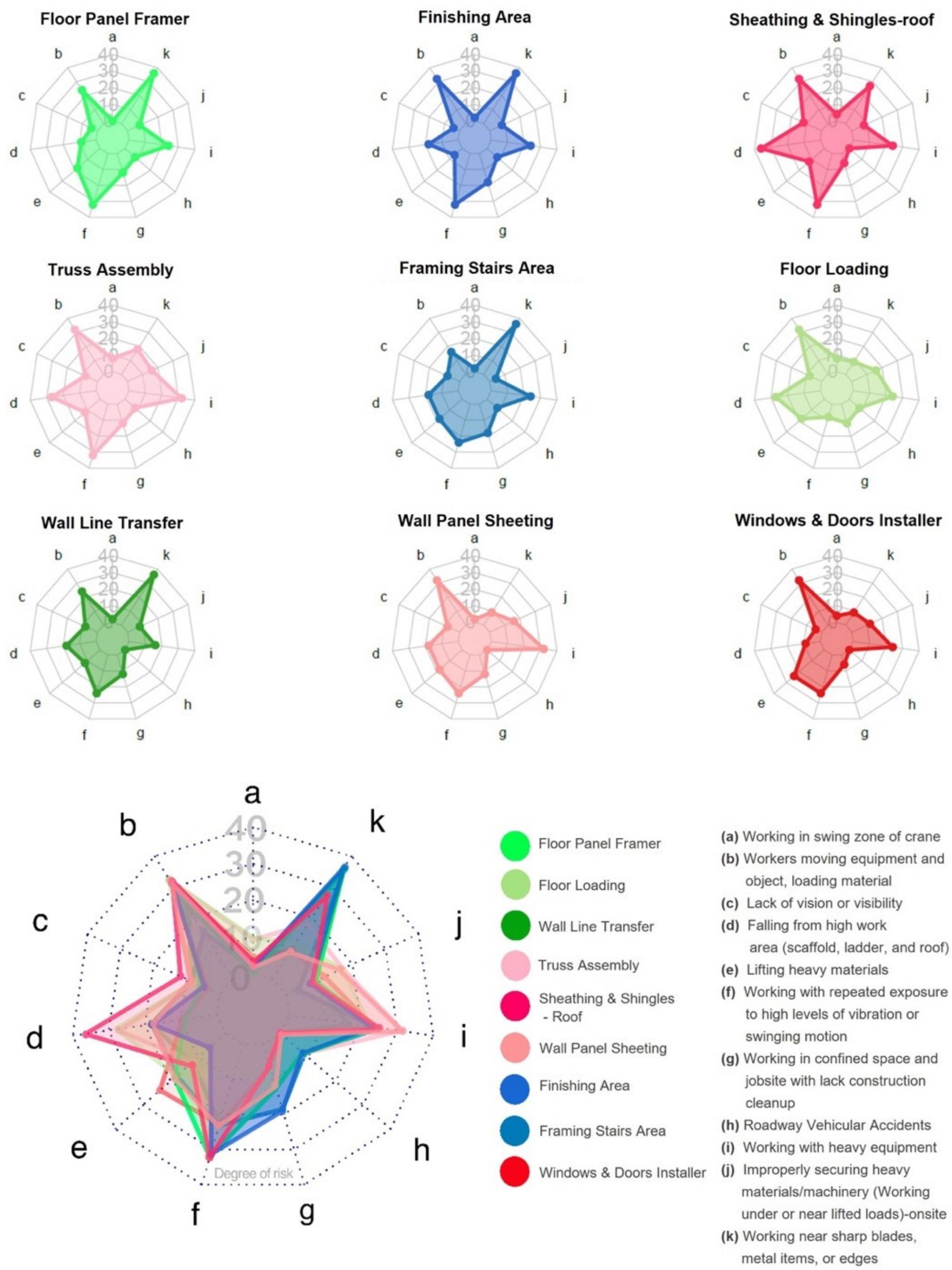


Figure 8: Detailed job-wide degree of risk radar chart

The number of severity values for each individual construction related job, namely number of low, medium, and high risk were then determined according to previous methodology using degree of risk analysis as shown in Table 4. The findings show that, with several high-risk attributes, "Finishing Area", "Sheathing and Shingles of Roof" and "Roof Truss Assembly" require the most significant investments among the nine job tasks. "Framing Stairs Area" and "Wall Line Transfer" have the highest number of medium risks followed by "Floor Loading", "Roof Truss Assembly" and "Windows & Doors Installation", which have the second highest number of medium risks. These number of classified risks tend to provide key risk perceptions and safety behaviours on specific job and the findings may aid project managers in their future work on construction site safety provisions. Once the number of injuries related to specific job is drastically increasing, company can suggest possible initiatives and put investment on reducing the load of this certain job by adjusting three factors (FE, IP and PC). All relevant degree of risk with detailed quantitative results as well as administrative and engineering control can then be provided based on key risk perceptions.

Table 4: Summarized number of severity values or risk classification for each job

Job Title	**Job Task Overview**	**Number of low risks(L)***	**Number of medium risks(M)**	**Number of high risks(H)**
Floor				
Floor Loading	Responsible for loading completed floor panels and securing to trailers for delivery	5	5	1
Floor Panel Framer	Responsible for obtaining the correct materials, placing studs in the correct position for framing, and installing sheathing plywood boards	5	4	2
Manual Build				

Finishing Area	Standing and building prefabricated wood building components using power tools and non-powered hand tools as per requirement of the component that is being manually built	5	3	3
Framing Stairs Area	Stick framing a stair cage unit	4	6	1
Roof				
Sheathing and Shingles - Roof	Responsible for sheathing rooftops and installing shingles, sidings, waterproof paper and cutting vents	4	4	3
Truss Assembly	Responsible for loading, sorting, and assembling the truss into the roof frame. Depending on the type of roof, installing drywalls for dividing roof and addition of Tyvek home wrap for sidings	3	5	3
Wall				
Wall Line Transfer Operator	Hammering nails, removing unwanted nails, controlling the butterfly table	4	6	1
Wall Panel Sheeting	Sheeting the Wall frame – studs with wall panels	6	4	1
Window and Door				
Windows & Doors Installer	Operating power tools, lifting, pushing, dragging, and pulling; vacuum lift and wood	4	5	2

**Low Risk-degree of risk score 1 to 9; Medium Risk-score 12 to 27; High Risk-score 32-64 (OHS 2011)*

4.3 Data Mining-based Analysis

In parallel with the attribute-based risk assessment, data mining method was also used to provide the correlation coefficients of important attributes. Table 5 was summarized based on this study's finding regarding the correlation matrix, the correlation measures the degree of association between two attributes and a more positive value for the correlation implies a positive tendency and association. Again, injuries associated with these two highly correlated attributes were observed to be closely relevant to the previous results and discussions. The same results were observed when considering safety attributes for “workers moving equipment, and object, loading material”, “working with repeated exposure to high levels of vibration or swinging motion”, “working with heavy equipment” and “working near sharp bladed, metal items or edges”. Attributes that relate to “nail gun”, “vibration” and “nailing” are coupled, and this again show consistency of the data mining results. Moreover, other than commonly correlated attributes, industries and administrators should also acknowledge the following associated correlation: wrist injury with trusses, wrist/hand injury with wall, back/arm injury with wrench and back/hand injury corresponds with working at heights. The findings indicate that one can utilize these correlated results to effectively achieve safety measurements and preventive actions to minimize the likelihood of potential hazardous occurrences.

Table 5: Matrix Visualization – Relative Injury Correlation Matrix with coefficient

First Correlated Attribute*	**Second Correlated Attribute**	**Correlation coefficient**
Equipment/Source of Injury = heavy object	Motion injury = lifting	0.749
Motion Injury = falling	Safety Hazard = Working from Height	0.730
Area of injury = wrist	Equipment source of injury = trusses	0.704

Area of injury = wrist/hand	Equipment source of injury = wall	0.704
Equipment source of injury = nail gun	Ergonomic risk factors = vibration	0.653
Equipment source of injury = falling object	Safety hazard = inadequate guards	0.580
Area of injury = back/arm	Equipment source of injury = wrench	0.572
Area of injury = back/hand	Safety hazard = working from height	0.572
Ergonomic risk factors = vibration	Motion injury = nailing	0.558
Ergonomic risk factors = F/P*	Motion injury = lifting	0.532

**One should note that the negative or inverse association for the correlation are not shown in the table. F/P can be interpreted as ergonomic risk factors that contain both forceful exertion work and poor posture.*

4.4 Adoption of Proposed Conceptual Digitalization Framework

With the implementation of the above methods and approaches, the proposed conceptual framework was illustrated as shown in Figure 4 using the safety related documents and blended analysis of the occupational safety hazards and risk assessment for future digitalization. The hierarchy was constructed comprising three criteria, each of which was further explained with easily comprehended design and visualization. In the first level of the hierarchy, in-plant/field job demand and incident report, PDA and SOP were collected among the parameters of criteria influencing the construction safety and risk assessments. The chosen safety related documents often give priority to risk evaluation and enable the key identification of both quantitative and qualitative injury data.

Once information is linked to a digital format, the second level of hierarchy provides an objective assessment of characterized factors based on each performance of the injury reports as well as key contributions from PDA and SOP. Although at the current stage, the study incorporates subjective judgement of experts as one of the inputs and considerations, but future system can be objective based on calculation from useful data extraction to reduce the inconsistency of judgments in terms of reliability. The main objective of this hierarchy is to establish a database of physical demand, operations, equipment, risky behaviours, ergonomic risks, frequencies of exposures, degrees of risk classifications and injury correlations, then to embed and employ the abovementioned blended analysis to present a robust method for prioritization of safety risks. Since objective assessment is used to ensure the robustness of the judgements on occupational safety hazard and risk, system will automatically conduct data updates allowing the traceability of modifications. The resulting degree of risk identification for each job, high risk task and injury correlation can be predicted and evaluated proactively, which also facilitates the development of constructive injury prediction models, control algorithms, corrective actions, and improved employee trainings.

Finally, the final level of hierarchy will emerge the proposed digital platform into a digital twin-driven design (Tao et al. 2018), where real-time interaction, validation and convergence can be shared between physical document-based space and information space. Digital twin in manufacturing field can be utilized to link the real and digital environments, resulting in a simulated reproduction of a product or procedure that reveals real-time information and execution of advanced control schemes. A digital twin-driven system can comprehensively improve the management by developing digital mapping of physical products and forming closed-loop feedback and optimization (Nikolakis et al. 2019). Similarly, the feature of introducing digital mapping of physical data generated in the use of safety related data can unlock the full industry value of construction operations with a seamless connection and automated update. The success connections among digital platform and data fusion will support the continuous data transmission and up-to-date realization of occupational hazards and risks. The final level of Figure 3 summarizes a preliminary proposed integration of both information system and data based structured digital-driven concept.

This framework will (1) systematically connect information flow from all different departments relying on the existing information systems and databases; (2) virtually collect and present the real-time H&S processes, including worker's performance and risk control, raw data collection and investigation, corrective actions and control schema; and (3) have the aforementioned sections integrated to support real-time decision making and simulate the future states for any changes or improvements to the system. The framework can also guide decision makers to assess complex construction situations based on knowledge of several key parameters provided in an up-to-date risk factor database. This proposed conceptual framework will help with the execution of advanced risk control schemes and capable of proactively reducing development of injury claims, occupational risk and WMSDS.

4.5 H&S Information Flow, Information Management and Control

4.5.1 H&S Information Flowchart

The proposed H&S information flowchart as shown in Figure 9 is validated in a local collaborative industry partner as a case study. The case study focuses on the information flow of the H&S section and conceptual digitalization and adoption. Two aspects of the study are completed in this chapter: (1) proposed H&S information flow and its process documentation; and (2) integrated conceptual digitalization framework with blended risk analysis and control to support the decision marking on the occupational safety hazard and risk assessment. In addition to that, the flowchart also introduces a knowledge-based decision-support model for H&S competence assessment of organizations, it help to support individual's decision-making and reasoning process for competence assessment.

This process documentation outlines (See Figure 9) the proposed activities, information flow that can be carried out by the H&S department in the construction enterprise. H&S activities are achieved through specific processes and these distinct processes are used to improve consistency of information flow in H&S hazard control. Processes are transparent, agile, and efficient and the proposed documentation elaborates how the H&S processes and information are implemented through detailed instructions, it is also intended to assist decision makers during the construction assignment of the impacts for prioritised risk factors on a digital platform, which can then be used in a digital twin-driven design that allows for future risk prediction and digital mapping of physical safety related data. The H&S process on the information flow empowers managers and other stakeholders by offering a top-level view of various processes. It's also handy as training material and for onboarding new H&S employees. Managers can spend less time teaching new employees the basics since they can turn to documentation to learn how a process is completed. Documentation also speeds up the learning curve, helping employees become productive quickly and can be viewed at as a blueprint for conducting H&S management activities.

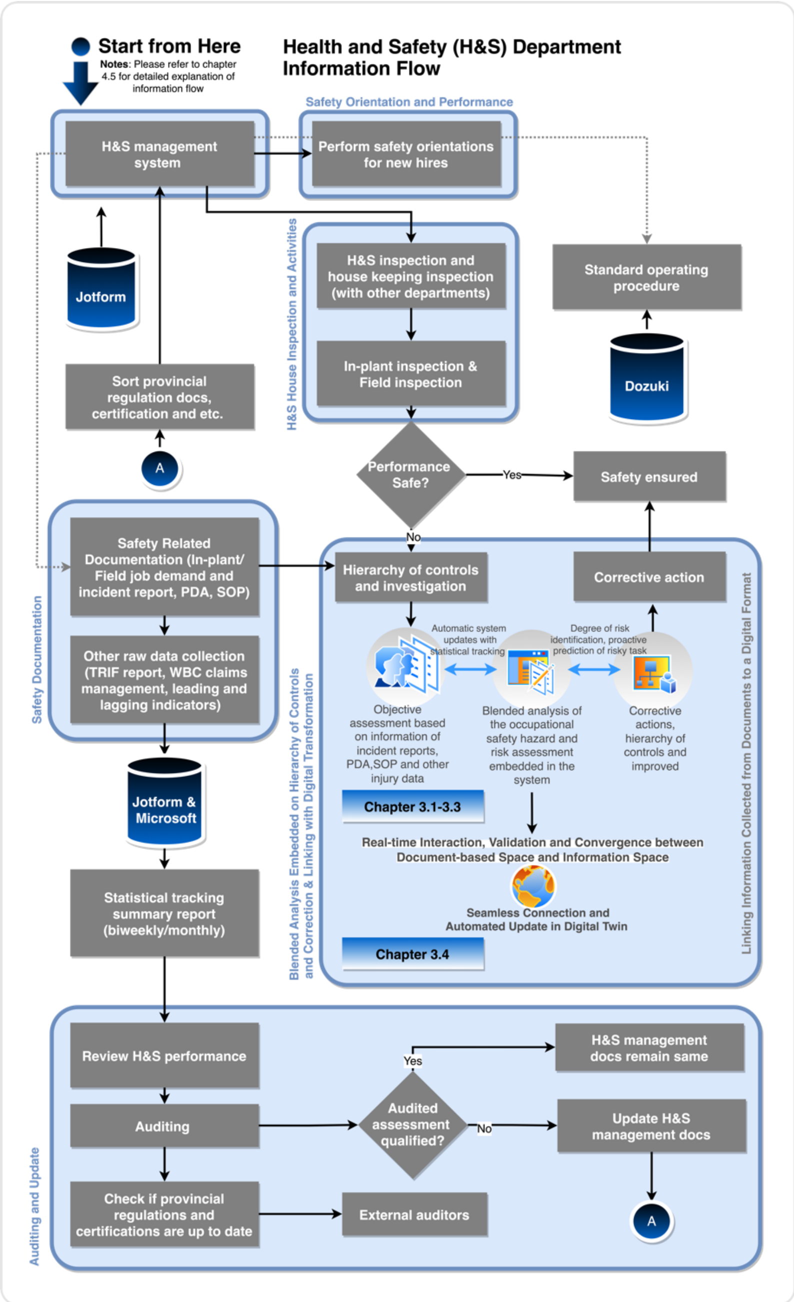


Figure 9: H&S Department Information Flow

The scope of this documentation is to propose a top-level overview of the H&S information flow that is carried out by the collaborative industry partner. This documentation will provide a complete description of the H&S and risk control activities and processes using words, images, and flow charts to demonstrate how the organization functions. The top-level overview of the business provides a macro view of the H&S activities demonstrating the overall flow and systemic features of the organization i.e., high-level view provides an executive summary of activities. Detailed description on each block diagram provides a micro view of the specific processes in the flowchart i.e., low-level view is concerned with individual components within the system and how they operate.

Effective H&S work and information flow is designed through a typical H&S management system (See Figure 9), the major activities and information flow that occur from this stage are "safety related documentation collection", "provincial regulation documentation and certification sort", "Health and safety inspection and housekeeping inspection", "Safety orientation performance" and standard operating procedure set up". Several other departments such as the "production department", "construction department", and the "logistics department" are collectively responsible for effective planning and data transferring.

The "safety related documentation collection" stage is first handled by H&S department and the resources of these raw H&S data, especially incident report, PDA, SOP, TRIF report, WBC claims, leading and lagging indicators, are acquired from workers, and converted into useful reports for higher level of investigation. These reports can be collected and updated through a cloud base report system and then a statistical tracking summary report can be generated on a monthly or weekly basis. In addition, all safety documentations will also feed into a hierarchy of controls and investigation in which will be discussed in more detailed later. Then followed by the H&S performance review, "auditing" stage will follow by experienced team and members. The duties of the H&S "auditing" are: 1) encouraging regular H&S system assessments to prevent them from deteriorating due to bad habits; 2) displaying a strong commitment to preserving the general H&S of workers; 3) ensuring that H&S management and personnel are correctly utilising safety programmes; 4) facilitating deliberate risk control's programming, policy, and process reforms and as well as assisting in the H&S resources division's areas for development; 5) assessing the performance of the management controls in place and minimising the chance of an accident or injury and providing a safe workplace for employees by evaluating their performance and effectiveness of the company's safety training in an objective way; and 6) confirming adherence to the relevant safety and health rules by locating and recognising potential dangers (existing and new). If the auditing assessment is qualified, H&S management documentation remain the same, conversely, an update to the H&S management is necessary through conjunction A in Figure 9 (includes other sorting on provincial documentations and certification through any online shared forms that could be fed into the H&S system. E.g., JotForm). Once auditing process has been targeted, employees are encouraged to check if provincial regulations and certification are up to date. External auditees will ensure and evaluate the effectives of the existing management controls.

The "Health and safety inspection and housekeeping inspection" stage consists of key H&S activities that facilitate the risk control function by enabling safety and health information flow. House inspection can be further subdivided into both in-plant inspection and field inspection for safe performance evaluation. The most important function and flow in this chart is the "hierarchy of controls and investigation", that is, objectively evaluating the adequacy of the company's safety performance and implementing corrective actions. This part is proposed based on the digital framework previously discussed in chapter 3.4 to assess digital mapping (Tao et al. 2018) and potential predictive metrics for efficient risk management and current risk factor assessment from chapter 3.1 to chapter 3.3. By linking the information collected from documentation to a digital form, the hierarchy of controls offers an objective evaluation of described criteria and factors based on each injury report's performance as well as significant contributions from PDA, SOP and other safety raw data. The above-mentioned blended analysis is then embedded and used to present a robust for prioritising safety risks. The subsequent degree of hazard identification for each job, significant risk task, and injury correlation can be predicted and examined proactively, enabling it to develop constructive injury forecasting model, adaptive control, corrective actions, and better safety assurance. After corrective actions has been developed, safety is ensured. These modules and processes are seamlessly integrated through a central database or digital-driven database in order to generate the digital mapping of physical data for current and future prediction and simulation. Noticed that H&S management system could also be liable to the activity of monitoring and controlling new hires safety orientation performance, therefore information flow through the safety orientations that can be assisted further by the standard operating procedure and training software. E.g., Dozuku. Employ hazard assessment data to determine what worker training is required, as well as to develop the content of employee orientations and job-specific training.

The core of H&S information flow involves holistic processes and invariably, process documentation allows the organization to track, understand, and analyze ongoing processes so that workers can replicate and improve it in the future. The H&S department manages a massive amount of information, particularly when it comes to information transfer and communication among various departments when human manual information processing is required. This documentation provides the opportunity to learn about each process and how it is carried out. The process documentation of the information flowchart can be beneficial for the industries in the following ways: 1) Improve and streamline various processes for H&S department; 2) Increase transparency across the organization; 3) Facilitates scaling the H&S management with high flexibility; 4) Make H&S and risk control knowledge transfer more convenient; 5) Analyze and identify flaws of H&S management processes; 6) Identify opportunities for continuous improvement; 7) Save time and labor power and 8) Reduce overall costs.

4.5.2 H&S Department Information Management and Control

The proposed H&S department information management and control is shown in Figure 10. As can be seen from Figure 10, three stages of activities are organized hierarchically following a systematized information management. For each engineering stage, the module demonstrates the high-level information management's target setting and tactics decision. The whole system is described at various level with focus on the "Management and Control for Construction Safety", "Information Management at Enterprise level" and "Information Management at Departmental level". For the occupational health, process safety, and work environment protection, strategy "Management and Control for Construction Safety" reflects the corporate safety values for the social demand. This makes it possible to specify the function of the overall integrated H&S information control, as the result of implementing all relevant embedded management activities. Each activity is adapted in the form of the current OHS standards with resource provision. "Information Management at Enterprise level" is then established to determine enterprise strategic

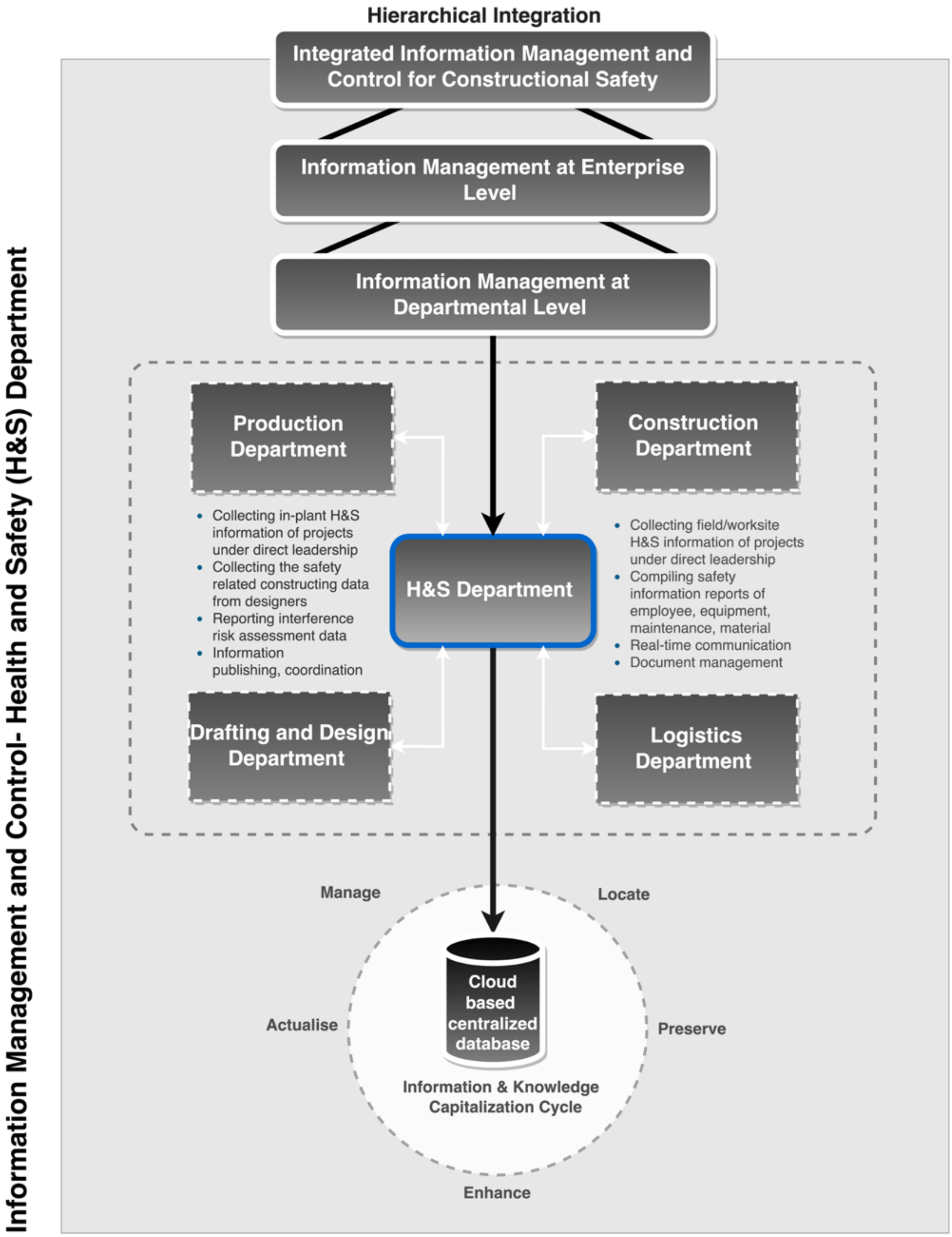


Figure 10: Framework of Proposed H&S Department Information Management and Control

plan, workflows, business information, and system function to be utilized in entire company's H&S information flow. Once the requirements of enterprise-level management are defined, it is necessary to determine the most opportune and practical technique of information monitoring, basing on specific "Information Management at Departmental level" for the H&S department deployment.

H&S department information management in Figure 10 adopts both upward reporting mechanism and downwards transmitting mechanism (Qi et al. 2009) through the interactions with other major departments. In this case, for example, the upward reporting mechanism is the pathway by which project-operation level managers collect all types of project H&S related information that occurs during the project constructing process, such as any physical injuries reports in workers that ranging from moderate to serious, any hazard identification and working conditions claims, interference risk assessment of the historical workplace, and input the basic data into the system. The information is then statistically collected and analysed by the intermediate functions layer; this could be a strategical embedded risk analysis system as mentioned earlier in the chapters. Finally, information that stored in H&S department will be reported to the proper decision-makers. Subsequently, the downward transmitting mechanism is used to distribute all types of management decisions and plans to the corresponding projects under direct leadership. Noticed that all information will deploy a real-time communication and transparent information publishing through any online share platform, such as cloud-based database system, online shared folders and etc.

Production, drafting and design, construction, logistics department are displayed as four most closely relevant departments for upward and downward data transmission and information management. Production department consists of construction activities which seek to allocate human resources, raw materials, and equipment/machines in a way that optimizes efficiency. It contains a various collection of in-plant H&S data. Similarly, drafting and design department prepares technical drawings and plans (3D model structural design documents, roof penalization plan, and building design) and this documentation often provides visual H&S guidelines on how to construct the end product. Moreover, the design team can also collaborate with other departments for interference safety assessment and reporting, thus a seamless connection becomes crucial between H&S and design department. Most on-site installation job is carried out by the construction department, and it collaborates closely with the logistics department to effectively implement and control the forward and reverse flow and storage of goods, construction using trucks, cranes, and trailers, therefore it included a diverse set of in-plant/ field H&S data and information.

The major features, which characterize the information flow departmental management, are in-depth data communication, cross-departmental interaction and connection which allow user to monitor the project performance over a certain period. Key data can be transformed into charts, curves, and tables by the database through risk assessment implementation. Data is reported through H&S department, comparing between projects, and executive reports. The procedures and the system can be continuously improved by transferring up-to-date data to achieve practicable construction engineering in OHS.

Finally, once the information connectivity mechanism is established, it is possible to quantify the particular effect of critical information for the construction H&S through the cloud based centralised database. The database was embedded in the management to encapsulate and evaluate the H&S performance's outcome as well as transferring tacit knowledge into explicit knowledge. The knowledge or information capitalisation cycle proposed by Grundstein et al. (2003) can thus be used to ensure a systematic approach to corporate data acquisition, creation, and management. Through five stages, the cycle represents the process of capitalisation of strategic knowledge, which is required to successfully carry out processes and activities of the company's core information evaluation. Five distinct stages include (Grundstein et al. 2003) 1) Locate: Create a detailed map of company knowledge (in terms of sources and topics) to allow for more effective management of its architecture, features, and value, 2) Preserve: Identify resources for corporate data preservation such as modelling, formalisation, and conservation tools, 3) Enhance: Increase the value of knowledge by putting it to work for the company's development and expansion through a controlled dissemination process, 4) Actualise: Knowledge is evaluated, standardised, and enriched with both user and external experience in order to remove and update old dataset and maintain coherence while continuously combining and integrating new knowledge and information in a dynamic and extended environment and 5) Manage: Establish policies and rules of interaction among the stages of the cycle to control activities that enable strategic knowledge capitalisation.

The main contribution of the above proposed framework is to develop a detailed information management with a specific focus on the H&S department, such as a hierarchical integration of safety data management at each level of the department, in order to provide efficient operation control along with the effective support for the enterprise decision-making. A combination of both H&S department flow diagram (See chapter 4.5.1) and departmental Information Management and Control could be used to create a comprehensive digital representation for a high-level management. This proposed Information Management and Control will integrate all the existing Information Technology tools and connect information from the closely relevant department (production, drafting and design, construction, logistics). Both case study results can accelerate the monitoring and assessing of performance safety and health management tasks within organizations. Both of these examples are focused on managing safety at the project-level. Together two proposed models will help enabling real-time data collection and multi-scenario simulation purposes.

CHAPTER 5: CONCLUSIONS

5.1 Research Conclusions

Complexity on construction hazard control and occupational risk management has been a major concern in modern construction industries. Specifically, the current industry suffers from: (1) lack of robust and effective method for risk factors analysis and quantification on injury related reports; (2) limited reliability of human perception when integrating digitalization with safety risk data; and (3) variations in the H&S information management and control resulting from massive information transferring.

In this way, existing limitations can be improved with respect to the three main objectives: (1) develop a blended analysis on occupational risk and hazard assessment using both attribute-based and data mining approaches; (2) propose a conceptual digitalization framework using both safety related documents and developed blended analysis; and (3) create a H&S information flowchart, as well as an information management and control framework for the H&S department. The proposed methodology and frameworks have been implemented in real case studies.

This thesis presents blended research on occupational risk assessment, in perspective with both attribute-based and data mining approaches. By applying the distinctive quantitative analysis on level of injuries and knowledge of construction hazards into over 180 incident reports in a case study with the support of both PDA and SOP documentations. This thesis first explores a strategy of degree of risk classification, safety attribute-based risk identification, and datamining analysis with injury correlation matrix.

By referring to the results, one can see that attributes, such as “workers moving equipment, and object, loading material”, “working with repeated exposure to high levels of vibration or swinging motion”, “working with heavy equipment”, “falling from high work area” and “working near sharp bladed, metal items or edges” have the highest degree of risk which can be interpreted as the high risky characteristics in cases of construction-related injuries. In terms of job-wide perspective, jobs that associated with “Finishing Area,” “Roof Sheathing and Shingles of Roof,” and “Roof Truss Assembly” require the most significant attentions during the construction. Correlation coefficients performed by data mining further justified the importance of these attributes. The risk evaluation and potential correlation inferred from this research contribute effective risk identifications and help identify potentially risky demand and construction hazards. In addition, previously there were no control strategies provided (Hallowell and Gambatese 2007; Esmaeili and Hallowell 2012) and the current method allocates control recommendations and performance evaluation based on the quantitative analysis. By establishing these controls, the organization can offer possible initiatives and reinvest in diminishing the load of specific job.

In addition, a proposed conceptual digitalization framework is presented in this thesis for effective management of safety risks and up-to-date risk factor evaluation. The framework combines safety related documentation and the blended analysis into a digital platform. The main contribution of the proposed framework is its ability to link and extract meaningful safety attributes and injury data. This can later be applied in digital twin-driven design to enable future risk prediction and digital mapping of physical safety data. Moreover, the quantitative findings illustrated several major safety and occupational health issues for construction industry to assist in a proactive risk evaluation.

To overcome the limitation of lack of supervised information control, this study also proposed an information flowchart for the H&S department, as well as an information management and control framework for the H&S department that adopts both upward reporting mechanism and downward transmitting mechanism through the interactions with other departments. The proposed method can be generalized for all construction-related activities, the development of this knowledge-based decision-support framework provides a scientific approach to assisting decision-maker in assessing the health and safety competence. The overall integrated construction information management can ultimately promote overall economic, safety, and social sustainable development by combining the above concepts and frameworks.

5.2 Research Contributions

The research outcomes have resulted in several research contributions:

- The blended study on the occupational risk assessment, in perspective with both attribute-based and data mining approaches is developed to help the construction industry identify relationships on causes of safety hazards, key safety attributes and ergonomic characteristics, as well as the extent of injuries with the corresponding risk controls and preventions. It is also expected to be used by all construction activities with its detailed job-wide, company-wide safety classification.
- The conceptual digitalization framework is proposed to help decision-makers assess complex construction situations by utilizing knowledge of several key parameters from an up-to-date digital twin-driven risk factor database. This proposed conceptual framework

will aid in the execution of advanced risk control schemes and will be capable of proactively reducing injury claims, occupational risk, and WMSDS.

- An information flowchart that specifically focus on H&S division, as well as an information management and control framework for the H&S department is created. Proposed H&S information flowchart introduces a knowledge-based decision support framework for H&S competence assessment of organizations, it helps to support individual's decision-making and reasoning process for competence assessment. The proposed approach adopts both upward reporting mechanism and downwards data transmitting mechanism that characterizes the information flow through departmental management, with in-depth data communication, cross-departmental interaction, and connection in which allow user to monitor the project performance over a certain period.

5.3 Future Works

Limitations on blended risk analysis for construction related injury may still exist in evaluating the level of risk classification. The current methodology introduced in this study classified frequency of exposure, incident probability, and potential consequence of loss into a scale of 1 to 4 and this can be further improved by quantifying risk index into a more comprehensive score. The study outcomes will benefit strategic risk data retrieval by enhancing the accuracy of blended analysis with weighted level of risk classification (Instead of integer division of score value). Further algorithm could be adopted by identifying the accuracy of the data mining-based model. The categorized risk factor assignments in the preprocessing stage are inherently limited by the occurrence of unseen, misspelled, and missing textual data. Applying automated content analysis technique can extract new fruitful information buried in the mass of textual data from each injury report. This new information can also be used for the further validation on the safety attribute relationship. In addition, 11 identified safety attributes that adopted from the Esmaeili and Hallowell model are not mutually exclusive and therefore can also be further converged, connected and evaluated together. This may provide practical means for practitioners to track and monitor interactions on construction sites as well as multi-objective decision making.

Moreover, the proposed future digitalization model with digital twin-driven system is expected to simulate what-if scenario in the virtual platform to support educated risk forecast, which may require longer implantation time and management. Hence, further research is needed to develop future configurations and scales. The proposed method can be generalized for construction-related activities based on its sufficient safety attributes and text-mining application of distinct construction-related job task. Together with its detailed job-wide, company-wide safety classification, the model is expected to be used by all construction activities although further validation needs to be placed. The blended research with the adoption of conceptual digitalization model is intended to help the construction industry recognize associations between hazard factors and ergonomic elements, as well as the extent of injuries with the risk controls and preventions corresponding to them. The proposed model needs more practical application to further demonstrate its validity; however, based on the current construction-related injury cases studies, the reliability of the risk attribute identification is validated, and results are aligned with current literatures (Hallowell and Gambatese 2007; Villanova and Prades 2014).

In addition, the proposed H&S information flowchart, as well as an information management and control framework for the H&S department may requires external auditing and practicality investigation. More case studies on improving on-site safety management and safety professionals' awareness of safety risks can also be further investigated. Both on-site safety management and safety professional's awareness of safety risks can be incorporated to evaluate the practicality of future digitalization. Future data collection could use advanced web technologies which contains standardized forms, the use of advanced web technologies also has the potential to improve knowledge acquisition, retrieval, and optimise health and safety management.

REFERENCES

Abdelhamid, S., Tariq, G., and Everett, J. 2000. Identifying root causes of construction accidents. Journal of Construction Engineering and Management, 126: 52–60. doi: 10.1061/(ASCE)0733-9364(2000) 126:1(52).

Aksorn, T., and Hadikusumo, B.H.W. 2008. Critical success factors influencing safety program performance in Thai construction projects. Safety Science, 46: 709–727. doi: 10.1016/j.ssci.2007.06.006.

Baradan, S., and Usmen, M.A. 2006. Comparative injury and fatality risk analysis of building trades. Journal of Construction Engineering and Management, 132(5): 533–539. doi:10.1061/(ASCE)0733-9364(2006)132:5(533).

Bo, D., and L., Zhouxue. 2002. "Discussion on Informationized Management of Building Enterprises from Viewpoint of Risk Control in Construction Group". Building Construction,Vol 24,No.6, pp.492~493

Basso, B., Carpegna,, C.D., Dibitonto,, C., Gaido,, G.K., and Zonato,, C.2004. Reviewing the safety management system by incident investigation and performance indicators. Journal of Loss Prevention in The Process Industries 17: 225–231. doi: 10.1016/J.JLP.2004.02.004

Brazier, A.J. 1993. A summary of incident reporting in the process industry. Journal of Loss Prevention Process Industries, 7(3): 243–248.doi:10.1016/0950-4230(94)80072-3.

Chi, S., and Han, S. 2013. Analyses of systems theory for construction accident prevention with specific reference to OSHA accident reports. International Journal of Project Management, 31: 1027–1041. doi: 10.1016/j.ijproman.2012.12.004.

Ciarapica, F.E. 2009. Classification and prediction of occupational injury risk using soft computing techniques: an Italian study. Safety Science,47(1): 36–49. doi: 10.1016/j.ssci.2008.01.006.

Cheung, S., K.K.W. Cheung, H.C.H. Suen, CSHM: web-based safety and health monitoring system for construction management, Journal of Safety Research 35 (2) (2004) 159–170.

Desvignes, M. Requisit. 2014. Emperical Risk Data for Integration of Safety with Advanced Technologies and Intelligent Systems. Master's Thesis, Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, CO, USA.

Dodge Data and Analytics 2017. Safety Management in the Construction Industry, Smart Market Report, Bedford, MA.

Esmaeili, B., and Hallowell, M. 2012. Attribute-based risk model for measuring safety risk of struck-by accidents. Construction Research Congress 2012: construction challenges in a flat world, ASCE, West Lafayette, IN, pp. 289–298.

Fang, D., Wu, C., and Wu, H. 2015. Impact of the supervisor on worker safety behavior in construction projects. Journal of Management in Engineering, 31: 04015001. doi:10.1061/(ASCE)ME.1943-5479.0000355.

Fayyad, U.M. 1996. Data mining and knowledge discovery: making sense out of data. IEEE Expert, 11(5): 20–25. doi:10.1109/64.539013.

Gagne, R. (n.d.), Developing a Legally Compliant Job Demands Analysis, March 2022. Available https://fit2wrk.com/wpcontent/uploads/2015/08/ARTICLE_Fit2wrk_ClinicalEd_vol1

Grieves, M, Digital Twin: Manufacturing Excellence Through Virtual Factory Replication, 2015. White Paper, Available https://www.researchgate.net/publication/275211047

Grundstein, M., Rosenthal-Sabroux, C. and Pachulski, A. (2003) 'Reinforcing decision aid by capitalizing on company's knowledge: future prospects', European Journal of Operational Research, Vol. 145, pp.256–272.

Government of Canada, Canadian Centre for Occupational Health and Safety (OHS). 2011. Hazard and risk: OHS answers-risk assessment. Canadian Centre for Occupational Health and Safety.

Hallowell, M.R., and Gambatese, J.A. 2007. A formal model of construction safety risk management. In Proceedings of the 2007 Construction and Building Research Conference, COBRA, sponsored by the Royal Institution of Chartered Surveyors and Georgia Tech, Atlanta, GA, 6-7September 2007.

Hallowell, M.R., Bhandari, S., and Alruqi, W. 2020. Methods of safety prediction: analysis and integration of risk assessment, leading indicators, precursor analysis, and safety climate. Construction Management and Economics, 38(4): 308–321. doi:10.1080/01446193.2019.1598566.

Haslam, R.A., Hide, S.A., Gibb, A.G.F., Gyi, D.E., Pavitt, T., Atkinson, S., and Duff, A.R. 2005. Contributing factors in construction accidents. Applied Ergonomics, 36: 401–415. doi: 10.1016/j.apergo.2004.12.002.PMID: 15892935.

Yassine, Q., "A Unified Digital Twin Framework for Real-time Monitoring and Evaluation of Smart Manufacturing Systems," 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), Vancouver, BC, Canada, 2019, pp. 1394-1401, doi:10.1109/COASE.2019.8843269.

Yin X., 2019. Application of BIM Technology in Architectural Design, Project Construction and Management [J]. Journal of Zhejiang Water Conservancy and Hydropower College, (02).

Kartam, N.A., Flood, I., and Koushki, P. 2000. Construction safety in Kuwait: issues, procedures, problems, and recommendations. Safety Science, 36: 163–184. doi:10.1016/S0925-7535(00)00041-2.

Kai, Z., “Study of Apply Strategy of Informationization in Construction Enterprise”, China Real Estate, 2004,Vol.280, No.4,pp.76~78

Khairunnisa, Latief, Y., Riantini, and Leni, S. 2020. Standard operational procedure (SOP) auditing process in integrated management system to improve the efficiency of organizational performance. IOP Conference Series: Earth and Environmental Science, 426(012018). doi:10.1088/1755-1315/426/1/012018.

Li, H., Al-Hussein, M., Lei, Z., and Ajweh, Z. 2013. Risk identification and assessment of modular construction utilizing fuzzy analytic hierarchy process (AHP) and simulation. Canadian

Journal of Civil Engineering,40(12): 1184–1195. doi: https://doi.org/10.1139/cjce-2013-0013.

Li, X., Fan, G., Abudan, A., Sukkarieh, M., Inyang, N. Gül, M., et al. 2015.Ergonomics and physical demand analysis in a construction manufacturing facility. In Proceedings of the 5th International/11th Construction Specialty Conference, Vancouver, BC, Canada, 8–10 June.

Li, X., Gul, M., and Al-Hussein, M. 2019. An improved physical demand analysis framework based on ergonomic risk assessment tools for the manufacturing industry. International Journal of Industrial Ergonomics,70: 58–69. doi: 10.1016/j.ergon.2019.01.004.

Liao, C.W., and Perng, Y.H. 2008. Data mining for occupational injuries in the Taiwan construction industry. Safety Science, 46: 1091–1102.doi: 10.1016/j.ssci.2007.04.007.

McCabe, B., Loughlin, C., Munteanu, R., Tucker, S., and Lam, A.L. 2008.Individual safety and health outcomes in the construction industry. Canadian Journal of Civil Engineering, 35(12): 1455–1467. doi: https://doi.org/10.1139/L08-091.

Mitropoulos, P., and Namboodiri, M. 2011. New method for measuring the safety risk of construction activities: task demand assessment. Journal of Construction Engineering and Management——ASCE,137(1): 30–38. doi:10.1061/(ASCE)CO.1943-7862.0000246.

National Institute for Occupational Safety and Health Office of the Director (NIOSH). 2007. Elements of ergonomic programs: use checklists. Centers for Disease Control and Prevention.

Nesmith, G., Keating, J.T., and Zacharias, L.A. 2013. Investigating process safety near misses to improve performance. American Institute of Chemical Engineers, 32(2).

Nikolakis,, N., Alexopoulos,, K., Xanthakis,, E., and Chryssolouris G. 2019. The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor. International Journal of Computer Integrated Manufacturing 32: 1 – 12. doi:10.1080/0951192X.2018.1529430

Prades, M. 2014. Attribute-based risk model for assessing risk to industrial construction tasks. Master's thesis, University of Colorado, Boulder, CO. R core Team.2014. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria

Paolo, P., E. Lazovik, J. Broekhuijsen, M. Borth and J. Verriet, "Lifecycle Governance for Effective Digital Twins: A Joint Systems Engineering and IT Perspective," 2020 IEEE International Systems Conference (SysCon), Montreal, QC, Canada, 2020, pp. 1-8, doi: 10.1109/SysCon47679.2020.9275662.

Qi, S., Zheng. L., Zhang. S., Ding. L., Luo. H. and Ke. H., "Study on Integrated and Controlling Management Information System of Construction Enterprise Groups," 2009 International Forum on Information Technology and Applications, 2009, pp. 425-428, doi: 10.1109/IFITA.2009.147.

Redinger, C., Dotson, K. and Leibowitz, A. (2011). Occupational Health and Safety Management Systems. In Patty's Industrial Hygiene, R. Harris (Ed.). https://doi.org/10.1002/0471435139.hyg049.pub2

Suraji, A., Duff, A.R., and Peckitt, S.J. 2001. Development of causal model of construction accident causation. Journal of Construction Engineering and Management, 127(4): 337–344. Doi:10.1061/(ASCE)0733-9364(2001)127:4(337)

Shimada, Yukiyasu, Teiji Kitajima, Tetsuo Fuchino and Kazuhiro Takeda. “An integrated management for occupational safety and health throughout the plant-lifecycle.” (2014)

Tam, C.M., Zeng, S.X., and Deng, Z.M. 2004. Identifying elements of poor construction safety management in China. Safety Science, 42: 569–586. doi:10.1016/j.ssci.2003.09.001.

Tao,, F., Cheng,, J., Qi, , Q., Zhang,, M., Zhang,, H., and Sui,, . 2018. Digital twin-driven product design, manufacturing and service with bigdata. The International Journal of Advanced Manufacturing Technology94: 3563–3576. doi:10.1007/S00170-017-0233-1

Teizer, J., Caldas, C.H., and Haas, C.T. 2007. Real-Time Three-Dimensional Occupancy Grid Modeling for the Detection and Tracking of Construction Resources. ASCE Journal of Construction Engineering and Management, 133, pp.880-888.

Thomson, B. 1996. Investigation of equipment related injuries, and fatalities in construction. MS thesis, University of Washington.

Tixier, A., Hallowell, M., Rajagopalan, B., and Bowman, D. 2016. Automated content analysis for construction safety: a natural language processing system to extract precursors and outcomes from unstructured injury reports. Automation in Construction, 62: 45–56.

Uzo, E., and Mohamed, H.I. 2018. Case study application of prevention through design to enhance workplace safety and health in Manitoba heavy construction projects. Canadian Journal of Civil Engineering46(2): 124–133. doi:10.1139/cjce-2017-0454.

Vladimirovna, N.E., & Federation., R. (2014). Occupational health and safety management system as a tool for ensuring safety of works in construction industry. Vestnik MGSU, 166-173.

Verma, A., Khan, S.D, Maiti, J., and Krishna, O.B. 2014. Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports. Safety Science, 70: 89–98.doi: 10.1016/j.ssci.2014.05.007.

Villanova, M.P. 2014. Attribute-based risk model for assessing risk to industrial construction tasks. MS thesis, University of Colorado.

Wang, H.H. 2013. Concept structure discovery using formal concept analysis. In Proceedings of the 30th ISARC, Montréal, Canada, pp. 905–913. ISBN 978-1-62993-294-1, ISSN 2413-5844. doi: https://doi.org/10.22260/ISARC2013/0098.

Winkel, G. Cocchio, J. Wilson, L. McCutcheon, D. 2003 n.d., A Handbook for Engineering Safety and Risk Management, Leadership in Risk Management, The University of Alberta Press 2003-08-15, Edmonton, Alberta, Canada.

Yi, Y. 2019. Development and implementation of safety information management system for construction enterprises. In Proceedings of the2019 International Conference on Smart Grid and Electrical Automation, ICSGEA, 2019, pp. 501–504.

Yildiz, A.E., Dikmen, I., and Birgonul, M.T. 2014. Using expert opinion for risk assessment: a case study of a construction project utilizing a risk mapping tool. Procedia Social and Behavioral Sciences, 119:519–528. doi: 10.1016/j.sbspro.2014.03.058

Yu, H. 2009. A knowledge-based system for construction health and safety competence assessment. University of Wolverhampton. United Kingdom.

Zweber, J.V., Kolonay RM, Kobryn P, Tuegel EJ. Digital Thread and Twin for Systems Engineering: Requirements to Design. 55th AIAA Aerospace Sciences Meeting: 2017.

Zhang, C., Xu, W., Liu, J., Liu, Z., Zhou., Z & Duc Truong Pham (2019) Digital twin-enabled reconfigurable modeling for smart manufacturing systems, International Journal of Computer Integrated Manufacturing, DOI: 10.1080/0951192X.2019.1699256

Zhou, W., Whyte, J. and Sacks, R. 2012. Construction safety and digital design: A review, Automation in Construction, 22, pp.102–111.

Zapalac, R., K. Kuemmler, and T. Malagon, “Establishing Management Information Systems for Multiproject Programes”. Journal of Management in Engineering,Vol. 10, No. 1, January/February 1994, pp. 37-42, (doi 10.1061/(ASCE)9742-597X

APPENDIX A

Detailed attribute categorization and key groupings

(Based on the pre-assigned categorized factors)

Summarized number of severity values or risk classification for each job

Safety Attributes	Corresponding Column and Keywords
Motion and Physical characteristics	
Working in swing zone of crane	motion injury (craning)
	equipment source of injury (crane)
Workers moving equipment and object, loading material	motion injury (carrying)
	motion injury (pulling)
	motion injury (pushing)
Lack of vision or visibility	safety hazard (Weather condition)
Falling from high work area (scaffold, ladder, and roof)	motion injury (falling)
	safety hazard (Working from Height)
Lifting heavy materials	motion injury (lifting)
	equipment source of injury (heavy object)
Working with repeated exposure to high levels of vibration or swinging motion	motion injury (nailing, swinging) ergonomic risk factors (vibration)
	equipment source of injury (nail gun)
	equipment source of injury (hammer)
Workstation and Jobsite	
Working in confined space and jobsite with lack of debris management	safety hazard (Inadequate clearances)

Roadway Vehicular Accidents/Possibility that a car or trailer intrudes into the work zone	motion injury (driving) equipment source of injury (vehicle)
Equipment, Material and Source	
Working with heavy equipment	equipment source of injury (heavy object)
Improperly securing heavy materials/machinery (Working under or near lifted loads)	safety hazard (Inadequate Guards and Protection) equipment source of injury (falling object)
Working near sharp blades, metal items, or edges	safety hazard (Inadequate PPE) equipment source of injury (metal item) equipment source of injury (foreign object)

R analysis coding (Data preprocessing and visualization)

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(opendatatoronto)
library(ggthemes)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library("pander")
library("kableExtra")
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```r
library("zoo")
```

```
##
## Attaching package: ' zoo'

## The following objects are masked from ' package:base' :
##
##     as.Date, as.Date.numeric
```

```r
library("lubridate")
```

```
##
## Attaching package: ' lubridate'

## The following objects are masked from ' package:base' :
##
##     date, intersect, setdiff, union
```

```r
daily_data <- "Injury Data Analysis.xlsx"
reported_raw <- read_excel(daily_data, sheet = 1) %>%
  clean_names()
```

```
## New names:
## * ` ` -> ...9
## * ` ` -> ...17
## * ` ` -> ...18
```

```r
reported_raw <- reported_raw[-c(15, 16, 22, 29, 30, 31, 39, 48, 59, 72, 83, 94, 181, 182,  183, 184, 185

reported_raw$date <- as.Date(as.numeric(reported_raw$date))

reported_raw$date <- reported_raw$date %m-% years(70) %m-% days(1)

reported_raw<-reported_raw %>%
  mutate(re_categorized_area = case_when(re_categorized_area=="plant A or plant B"~"Both",
                              re_categorized_area=="both"~"Both",
                              re_categorized_area=="field"~"field",
                              re_categorized_area=="logistics"~"logistics",
                              re_categorized_area=="Both"~"Both",
                              re_categorized_area=="plant A"~"plant A",
                              re_categorized_area=="plant B"~"plant B"))

data <- reported_raw %>%
  select(motion_injury,
         area_of_injury,
         ergonomic_risk_factors,
         safety_hazard,
         equipment_source_of_injury)

reported_raw<-reported_raw %>%
```

```r
  utate(motion_injury = case_when(motion_injury=="carrying"~"carrying",
                                  motion_injury=="driving"~"driving",
                                  motion_injury=="exiting"~"exiting",
                                  motion_injury=="carrying"~"carrying",
                                  motion_injury=="falling"~"falling",
                                  motion_injury=="kneeling"~"kneeling",
                                  motion_injury=="lifting"~"lifting",
                                  motion_injury=="nailing"~"nailing",
                                  motion_injury=="pulling"~"pulling",
                                  motion_injury=="pushing"~"pushing",
                                  motion_injury=="slipping"~"slipping",
                                  motion_injury=="Slipping"~"slipping",
                                  motion_injury=="slippping"~"slipping",
                                  motion_injury=="swinging"~"swinging",
                                  motion_injury=="walking"~"walking",
                                  motion_injury=="craning"~"craning"))



reported_raw<-reported_raw %>%
  mutate(area_of_injury = case_when(area_of_injury=="Abdomen"~"chest",
                                    area_of_injury=="ankle"~"Ankle",
                                    area_of_injury=="arm and legs"~"leg",
                                    area_of_injury=="arm/hand"~"arm",
                                    area_of_injury=="arm/shoudler"~"arm",
                                    area_of_injury=="back/arm"~"back",
                                    area_of_injury=="back/hand"~"back",
                                    area_of_injury=="back/shoulder"~"shoulder",
                                    area_of_injury=="Finger"~"finger",
                                    area_of_injury=="wrist/hand"~"wrist",
                                    area_of_injury=="Lower Back"~"back",
                                    area_of_injury=="back"~"back",
                                    area_of_injury=="Knee"~"knee",
                                    area_of_injury=="wrist/hand"~"wrist",
                                    area_of_injury=="wrist/hand"~"wrist",
                                    area_of_injury=="hand"~"hand",
                                    area_of_injury=="knee"~"knee",
                                    area_of_injury=="finger"~"finger",
                              area_of_injury=="wrist"~"wrist",
                              area_of_injury=="shoulder"~"shoulder",
                              area_of_injury=="Shoulder"~"shoulder",
                              area_of_injury=="arm"~"arm",
                              area_of_injury=="foot"~"foot",
                              area_of_injury=="head"~"head",
                              area_of_injury=="face"~"face",
                              area_of_injury=="finger"~"finger",
                              area_of_injury=="Hand"~"hand",
                              area_of_injury=="Foot"~"foot",
                              area_of_injury=="Ankle"~"Ankle",
                              area_of_injury=="hip"~"hip",
                              area_of_injury=="stomach"~"chest",
                              area_of_injury=="hand/finger"~"hand",
                              area_of_injury=="hand/finger"~"hand",
```

```r
                         area_of_injury=="Neck"~"head",
                         area_of_injury=="neck"~"head",
                         area_of_injury=="chest"~"chest"))

reported_raw<-reported_raw %>%
  mutate(ergonomic_risk_factors = case_when(ergonomic_risk_factors=="F/P"~"F/P",
                                           ergonomic_risk_factors=="R/P"~"R/P",
                                               ergonomic_risk_factors=="Forceful Exertions"~"Forceful
                                               ergonomic_risk_factors=="poor posture"~"poor posture",
                                               ergonomic_risk_factors=="repetiton"~"repetiton",
                                               ergonomic_risk_factors=="vibration"~"vibration",
                                               ergonomic_risk_factors=="Vibration"~"vibration"))


reported_raw<-reported_raw %>%
  mutate(safety_hazard = case_when(safety_hazard=="inadequate guards"~"Inadequate Guards and Protection
                                   safety_hazard=="inadequate clearances"~"Inadequate clearances",
                                  safety_hazard=="Inadequate clearances"~"Inadequate clearances",
                                  safety_hazard=="inadequate PPE"~"Inadequate PPE",
                                  safety_hazard=="weather condition"~"Weather condition",
                                  safety_hazard=="Weather condition"~"Weather condition",
                                  safety_hazard=="working from height"~"Working from Height",
                                  safety_hazard=="Working from Height"~"Working from Height",
                                  safety_hazard=="inadequate maintenance"~" Inadequate maintenance",
                                  safety_hazard=="inadequate equipment and tool "~"Inadequate
                                  Equipmentsafety hazard=="Inadequate training  "~" Inadequate
                                  training"))

reported_raw<-reported_raw %>%
  mutate(equipment_source_of_injury = case_when(equipment_source_of_injury=="crane"~"crane",
                                   equipment_source_of_injury=="falling object"~"falling object",
                                  equipment_source_of_injury=="falling objects(panel)"~"falling objec
                                  equipment_source_of_injury=="foreign object"~"foreign object",
                                  equipment_source_of_injury=="Foreign Object"~"foreign object",
                                  equipment_source_of_injury=="heavy objects(truss)"~"heavy object",
                                  equipment_source_of_injury=="machine"~"machine",
                                  equipment_source_of_injury=="machines"~"machine",
                                  equipment_source_of_injury=="nail gun"~"nail gun",
                                  equipment_source_of_injury=="truss"~"heavy object",
                                  equipment_source_of_injury=="Exavator"~"vehicle",
                                  equipment_source_of_injury=="falling objects(wall)"~"falling object
                                  equipment_source_of_injury=="hammer"~"hammer",
                                  equipment_source_of_injury=="ice"~"ice",
                                  equipment_source_of_injury=="metal item"~"metal item",
                                  equipment_source_of_injury=="propane"~"propane",
                                  equipment_source_of_injury=="trusses"~"heavy object",
                                  equipment_source_of_injury=="falling objects(window)"~"falling obje
                             equipment_source_of_injury=="heavy object"~"heavy object",
                           equipment_source_of_injury=="joist"~"heavy object",
                           equipment_source_of_injury=="metal items"~"metal item",
                           equipment_source_of_injury=="roof"~"heavy object",
                           equipment_source_of_injury=="vehicle"~"vehicle",
                             equipment_source_of_injury=="falling object(wall)/trailer"~"falling objec
                           equipment_source_of_injury=="heavy objects(window)"~"heavy object",
```

```
                              equipment_source_of_injury=="ladder"~"ladder",
                              equipment_source_of_injury=="mud"~"mud",
                            equipment_source_of_injury=="trailer"~"trailer",
                              equipment_source_of_injury=="wall"~"heavy object",
                              equipment_source_of_injury=="falling objects"~"falling object",
                            equipment_source_of_injury=="Foreign object"~"foreign object",
                            equipment_source_of_injury=="heavy objects"~"heavy object",
                              equipment_source_of_injury=="trailer/falling objects"~"falling object",
                              equipment_source_of_injury=="wrench"~"wrench"))

reported_raw<-reported_raw %>%
  mutate(type = case_when(type=="First Aid"~"First Aid",
                                    type=="FIRST AID"~"First Aid",
                                   type=="Lost Time"~"Lost Time",
                                   type=="Medical Aid"~"Medical Aid",
                                   type=="Near Miss"~"Near Miss",
                                   type=="NEAR MISS"~"Near Miss",
                                    type=="PROPERTY DAMAGE"~"Property Damage",
                                   type=="Property  Damage"~"Property  Damage",
                                   type=="Statement"~"Near Miss"))

write.csv(data, "final.csv")

new_su <- reported_raw %>%
  select(motion_injury,
         area_of_injury)

new_su  <- na.omit(new_su)

new_su2 <- reported_raw %>%
  select(area_of_injury,
         safety_hazard)
new_su2 <- na.omit(new_su2)

new_su3 <- reported_raw %>%
  select(area_of_injury,
         ergonomic_risk_factors)
new_su3 <- na.omit(new_su3)

new_su4 <- reported_raw %>%
  select(area_of_injury,
         equipment_source_of_injury)
new_su4 <- na.omit(new_su4)

c <- ggplot(new_su, aes(x=motion_injury, fill=area_of_injury)) +
  geom_bar( ) +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1))+
  geom_text(stat='count', aes(label=..count..), position = position_stack(vjust= 0.5), size=3)+
  xlab("motion injury") +
```

```
kable(table(new_su$motion_injury, new_su$area_of_injury)) %>%
  column_spec(1, width = "5em")

b <- ggplot(new_su2, aes(x=safety_hazard, fill=area_of_injury)) +
  geom_bar( ) +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1)) +
  geom_text(stat=' count' , aes(label=..count..), position = position_stack(vjust= 0.5), size=3)+
  xlab("safety hazard") +
  ylab("number of cases")

kable(table(new_su2$safety_hazard, new_su2$area_of_injury)) %>%column_spec(1,
  width = "8em")

a <- ggplot(new_su3, aes(x=ergonomic_risk_factors, fill=area_of_injury)) +
  geom_bar(position = "stack") +
  geom_text(stat=' count' , aes(label=..count..), position = position_stack(vjust= 0.5), size=3)+
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1))+
  xlab("ergonomic risk factors") +
  ylab("number of cases")

kable(table(new_su3$ergonomic_risk_factors, new_su3$area_of_injury))%>%column_spec(1,
  width = "8em")

d <- ggplot(new_su4, aes(x=equipment_source_of_injury, fill=area_of_injury)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1))+
  geom_text(stat=' count' , aes(label=..count..), position = position_stack(vjust= 0.5), size=3)+
  xlab("equipment source of injury") +
  ylab("number of cases")
d

a <- as.matrix(table(new_su4$equipment_source_of_injury, new_su4$area_of_injury))a <-
as.data.frame(a)

  new_su5 <- data %>% select(motion_injury,
           area_of_injury, ergonomic_risk_factors)

  new_su5 <- na.omit(new_su5)

  level1 <- c("shoulder", "arm", "back", "finger", "hand", "foot", "head", "Ankle", "knee",level2 <-
  c("wrist")

  new_su5 <- new_su5 %>% mutate(area_by_hand = case_when(area_of_injury %in%
    level1~0,
    area_of_injury %in% level2~1))
```

```
## Call:
## glm(formula = area_by_hand ~ motion_injury + ergonomic_risk_factors,
##     family = "binomial", data = new_su5)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -2.409e-06  -2.409e-06  -2.409e-06  -2.409e-06  -2.409e-06
##
## Coefficients:
##                                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                               -2.657e+01  2.586e+05       0        1
## motion_injurykneeling                      7.035e-29  6.401e+05       0        1
## motion_injurylifting                      -7.487e-29  2.227e+05       0        1
## motion_injuryN/A                          -1.003e-29  2.368e+05       0        1
## motion_injurynailing                      -4.504e-29  2.547e+05       0        1
## motion_injurypulling                      -3.698e-29  3.950e+05       0        1
## motion_injurypushing                      -4.602e-29  6.401e+05       0        1
## motion_injuryslipping                     -2.569e-29  3.950e+05       0        1
## motion_injuryswinging                     -1.044e-29  2.496e+05       0        1
## motion_injurywalking                       8.573e-29  3.950e+05       0        1
## ergonomic_risk_factorsForceful Exertions  4.204e-30  2.350e+05       0        1
## ergonomic_risk_factorspoor posture         9.183e-30  2.441e+05       0        1
## ergonomic_risk_factorsR/P                  2.095e-29  5.597e+05       0        1
## ergonomic_risk_factorsrepetiton           -2.801e-29  3.243e+05       0        1
## ergonomic_risk_factorsvibration            2.263e-29  2.603e+05       0        1
## ergonomic_risk_factorsVibration           -2.083e-13  4.164e+05       0        1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.0000e+00  on 37  degrees of freedom
## Residual deviance: 2.2046e-10  on 22  degrees of freedom
##   (19 observations deleted due to missingness)
## AIC: 32
##
## Number of Fisher Scoring iterations: 25
```

```
factor(new_su5$area_of_injury)
```

```
##  [1] hand          back          back          hand          head
##  [6] foot          back          back          shoulder      N/A
## [11] arm/back      knee          arm           finger        arm
## [16] N/A           wrist/hand    hand/finger   chest         hand
## [21] finger        neck          finger        finger        hip
## [26] back          finger        back          shoulder      shoulder
## [31] finger        hand          back/shoulder finger        finger
## [36] wrist/hand    back/arm      back          hand          N/A
## [41] knee          arm/shoudler  knee          arm/hand      back
## [46] finger        finger        finger        hand          Lower Back
## [51] Lower Back    Ankle         Lower Back    Finger        Hand
## [56] Hand          Lower Back
## 23 Levels: Ankle arm arm/back arm/hand arm/shoudler back ... wrist/hand
```

```
## [1] ″2019-01-01″ ″2019-01-02″ ″2019-01-03″ ″2019-01-04″ ″2019-01-05″
## [6] ″2019-01-06″ ″2019-01-07″ ″2019-01-08″ ″2019-01-09″ ″2019-01-10″
## [11] ″2019-01-11″ ″2019-01-12″ ″2019-01-13″ ″2019-01-14″ ″2019-01-15″
## [16] ″2019-01-16″ ″2019-01-17″ ″2019-01-18″ ″2019-01-19″ ″2019-01-20″
## [21] ″2019-01-21″ ″2019-01-22″ ″2019-01-23″ ″2019-01-24″ ″2019-01-25″
## [26] ″2019-01-26″ ″2019-01-27″ ″2019-01-28″ ″2019-01-29″ ″2019-01-30″
## [31] ″2019-01-31″

## [1] ″2019-02-01″ ″2019-02-02″ ″2019-02-03″ ″2019-02-04″ ″2019-02-05″
## [6] ″2019-02-06″ ″2019-02-07″ ″2019-02-08″ ″2019-02-09″ ″2019-02-10″
## [11] ″2019-02-11″ ″2019-02-12″ ″2019-02-13″ ″2019-02-14″ ″2019-02-15″
## [16] ″2019-02-16″ ″2019-02-17″ ″2019-02-18″ ″2019-02-19″ ″2019-02-20″
## [21] ″2019-02-21″ ″2019-02-22″ ″2019-02-23″ ″2019-02-24″ ″2019-02-25″
## [26] ″2019-02-26″ ″2019-02-27″ ″2019-02-28″

## [1] ″2019-03-01″ ″2019-03-02″ ″2019-03-03″ ″2019-03-04″ ″2019-03-05″
## [6] ″2019-03-06″ ″2019-03-07″ ″2019-03-08″ ″2019-03-09″ ″2019-03-10″
## [11] ″2019-03-11″ ″2019-03-12″ ″2019-03-13″ ″2019-03-14″ ″2019-03-15″
## [16] ″2019-03-16″ ″2019-03-17″ ″2019-03-18″ ″2019-03-19″ ″2019-03-20″
## [21] ″2019-03-21″ ″2019-03-22″ ″2019-03-23″ ″2019-03-24″ ″2019-03-25″
## [26] ″2019-03-26″ ″2019-03-27″ ″2019-03-28″ ″2019-03-29″ ″2019-03-30″
## [31] ″2019-03-31″

## [1] ″2019-04-01″ ″2019-04-02″ ″2019-04-03″ ″2019-04-04″ ″2019-04-05″
## [6] ″2019-04-06″ ″2019-04-07″ ″2019-04-08″ ″2019-04-09″ ″2019-04-10″
## [11] ″2019-04-11″ ″2019-04-12″ ″2019-04-13″ ″2019-04-14″ ″2019-04-15″
## [16] ″2019-04-16″ ″2019-04-17″ ″2019-04-18″ ″2019-04-19″ ″2019-04-20″
## [21] ″2019-04-21″ ″2019-04-22″ ″2019-04-23″ ″2019-04-24″ ″2019-04-25″
## [26] ″2019-04-26″ ″2019-04-27″ ″2019-04-28″ ″2019-04-29″ ″2019-04-30″

## [1] ″2019-05-01″ ″2019-05-02″ ″2019-05-03″ ″2019-05-04″ ″2019-05-05″
## [6] ″2019-05-06″ ″2019-05-07″ ″2019-05-08″ ″2019-05-09″ ″2019-05-10″
## [11] ″2019-05-11″ ″2019-05-12″ ″2019-05-13″ ″2019-05-14″ ″2019-05-15″
## [16] ″2019-05-16″ ″2019-05-17″ ″2019-05-18″ ″2019-05-19″ ″2019-05-20″
## [21] ″2019-05-21″ ″2019-05-22″ ″2019-05-23″ ″2019-05-24″ ″2019-05-25″
## [26] ″2019-05-26″ ″2019-05-27″ ″2019-05-28″ ″2019-05-29″ ″2019-05-30″
## [31] ″2019-05-31″

## [1] ″2019-06-01″ ″2019-06-02″ ″2019-06-03″ ″2019-06-04″ ″2019-06-05″
## [6] ″2019-06-06″ ″2019-06-07″ ″2019-06-08″ ″2019-06-09″ ″2019-06-10″
## [11] ″2019-06-11″ ″2019-06-12″ ″2019-06-13″ ″2019-06-14″ ″2019-06-15″
## [16] ″2019-06-16″ ″2019-06-17″ ″2019-06-18″ ″2019-06-19″ ″2019-06-20″
## [21] ″2019-06-21″ ″2019-06-22″ ″2019-06-23″ ″2019-06-24″ ″2019-06-25″
## [26] ″2019-06-26″ ″2019-06-27″ ″2019-06-28″ ″2019-06-29″ ″2019-06-30″

## [1] ″2019-07-01″ ″2019-07-02″ ″2019-07-03″ ″2019-07-04″ ″2019-07-05″
```

```r
reported_raw  <-  reported_raw  %>% mutate(date_new  =  case_when(  date  %in%
  level1~"Jan 2019",
  date  %in%  level2~"Feb  2019",  date  %in%
  level3~"Mar  2019",  date  %in%  level4~"Apr
  2019",  date  %in%  level5~"May  2019",  date
  %in% level6~"Jun 2019",


  date %in% level7~"Jul 2019",  date %in%
  level8~"Aug 2019",  date %in% level9~"Sep
  2019",  date %in% level10~"Oct 2019",date
  %in% level11~"Nov 2019",date %in%
  level12~"Dec 2019",date %in% level13~"Jan
  2020",date %in% level4~"Feb 2020",  date
  %in% level15~"Mar 2020",date %in%
  level16~"Apr 2020",date %in% level17~"May
  2020",date %in% level18~"Jun 2020",date
  %in% level19~"Jul 2020"))


nnn <- reported_raw %>%select(date_new,
         motion_injury,  equipment_source_of_injury)

nnn  <-  na.omit(nnn)
write.csv(nnn,  "Working  in  swing  zone  of  crane_Incident  Probability.csv")


uuu <- reported_raw %>%
  select(date_new,
         motion_injury)

uuu  <-  na.omit(uuu)
write.csv(uuu,  "Workers  moving  equipment  and  object&Lifting  heavy  materials_Incident  Probability.csv")


yyy <- reported_raw %>%
  select(date_new,
         safety_hazard)

yyy  <-  na.omit(yyy)
write.csv(yyy,  "Lack  of  vision  or  visibility&Working  in  confined  space_Incident  Probability.csv")


qqq <- reported_raw %>%select(date_new,
         motion_injury,safety_hazard)

qqq  <-  na.omit(qqq)
write.csv(qqq,  "Falling  from  high  work_Incident  Probability.csv")
```

```r
ooo <- reported_raw %>%select(date_new,
        motion_injury, ergonomic_risk_factors,
        equipment_source_of_injury)

ooo <- na.omit(ooo)
write.csv(ooo, "Lifting heavy materials_Incident Probability.csv")


lll <- reported_raw %>%select(date_new,
        motion_injury, equipment_source_of_injury)

lll <- na.omit(lll)
write.csv(lll, "Roadway Vehicular Accidents_Incident Probability.csv")


kkk <- reported_raw %>%select(date_new,
        equipment_source_of_injury)

kkk <- na.omit(kkk)
write.csv(kkk, "Working with heavy equipment_Incident Probability.csv")


mmm <- reported_raw %>%
  select(date_new,
         equipment_source_of_injury,
        safety_hazard)

mmm <- na.omit(mmm)
write.csv(mmm, "Improperly securing heavy&Working near sharp blades_Incident Probability.csv")



uuu1 <- reported_raw %>%
  select(type,
        motion_injury)

uuu1 <- na.omit(uuu1)
uuu1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "P
               arrange(type) -> uuu1
write.csv(uuu1, "2_pot.csv")

 <- reported_raw %>%
  select(type,
         motion_injury,
         equipment_source_of_injury)

zzz <- na.omit(zzz)

zzz %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "Pr
               arrange(type) -> zzz
```

```r
qqq1 <- reported_raw %>%
  select(type,
          motion_injury,
         safety_hazard)

qqq1 <- na.omit(qqq1)
qqq1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "P
                arrange(type) -> qqq1
write.csv(qqq1, "4_pot.csv")

uu1 <- reported_raw %>%
  select(type,
         motion_injury)

uu1 <- na.omit(uu1)
uu1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "Pr
                arrange(type) -> uu1

write.csv(uu1, "5_pot.csv")

ooo1 <- reported_raw %>%
  select(type,
         motion_injury,
         ergonomic_risk_factors,
         equipment_source_of_injury)

ooo1 <- na.omit(ooo1)
ooo1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "P
                arrange(type) -> ooo1
write.csv(ooo1, "6_pot.csv")

yy1 <- reported_raw %>%
  select(type,
         safety_hazard)

yy1 <- na.omit(yy1)
yyy1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "P
                arrange(type) -> yy1
write.csv(yy1, "7_pot.csv")

1111 <- reported_raw %>%
  select(type,
         motion_injury,
         equipment_source_of_injury)

1111 <- na.omit(1111)
1111 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "P
                arrange(type) -> 1111
write.csv(1111, "8_pot.csv")
```

```r
kkk1 <- reported_raw %>%
  select(type,
         equipment_source_of_injury)

kkk1 <- na.omit(kkk1)
kkk1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "P
                 arrange(type) -> kkk1
write.csv(kkk1, "9_pot.csv")
```

```r
mmm1 <- reported_raw %>%
  select(type,
          equipment_source_of_injury,
         safety_hazard)

mmm1 <- na.omit(mmm1)
mmm1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "P
                 arrange(type) -> mmm1
write.csv(mmm1, "10_pot.csv")
```

```r
mm1 <- reported_raw %>%
  select(type,
          equipment_source_of_injury,
         safety_hazard)

mm1 <- na.omit(mm1)
mm1 %>% mutate(type = factor(type, levels = c("First Aid", "Lost Time", "Medical Aid", "Near Miss", "Pr
                 arrange(type) -> mm1
write.csv(mm1, "11_pot.csv")
```

```r
kkkk <- reported_raw %>%
  select(date,
         equipment_source_of_injury)

kkkk <- na.omit(kkkk)
kkkk[order(as.Date(kkkk$date, format="%d/%m/%Y")),]
```

```
## # A tibble: 122 x 2
##    date       equipment_source_of_injury
##    <date>     <chr>
##  1 2019-01-08 machine
##  2 2019-01-15 metal item
##  3 2019-01-16 machine
##  4 2019-01-17 heavy object
##  5 2019-01-19 vehicle
##  6 2019-01-19 vehicle
##  7 2019-01-21 metal item
##  8 2019-01-30 falling object
##  9 2019-01-31 nail gun
## 10 2019-02-21 vehicle
## # ... with 112 more rows
```

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(readxl)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```r
library(opendatatoronto)
library(ggthemes)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
library("zoo")
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library("lubridate")
```

```
##
## Attaching package: ' lubridate'

## The following objects are masked from ' package:base' :
##
##     date, intersect, setdiff, union
```

```
daily_data <- "Injury Data Analysis.xlsx"
reported_raw <- read_excel(daily_data, sheet = 1) %>%
  clean_names()
```

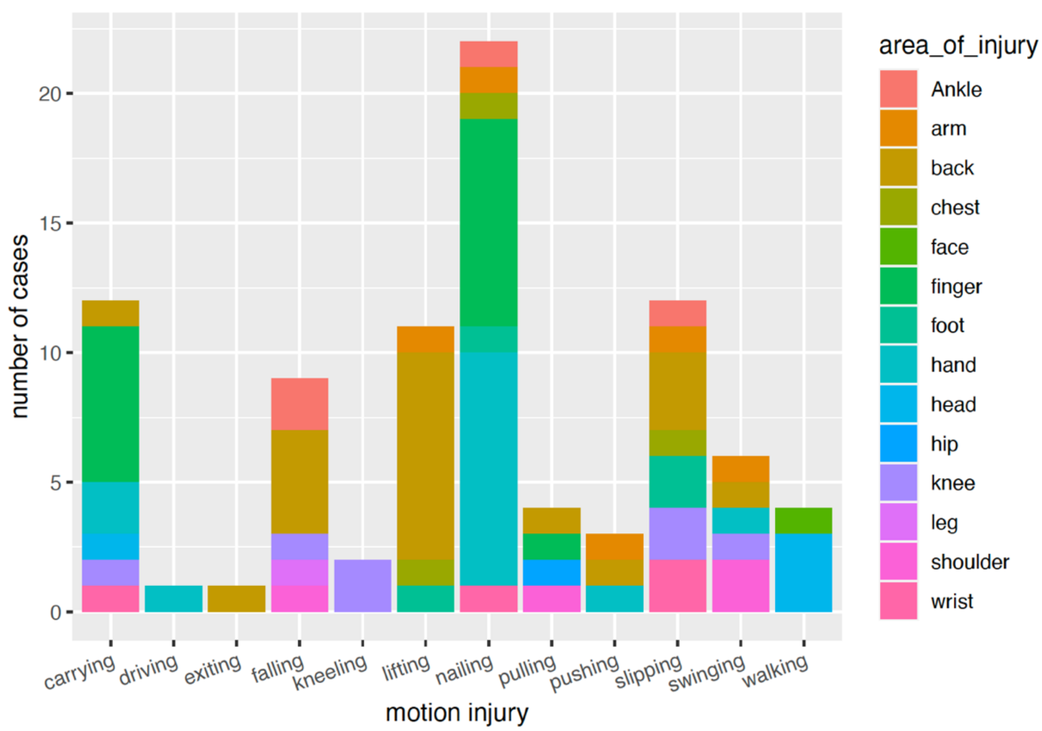
```
## New names:
## * ` ` -> ...9
## * ` ` -> ...17
## * ` ` -> ...18
```

```
reported_raw <- reported_raw[-c(15, 16, 22, 29, 30, 31, 39, 48, 59, 72, 83, 94, 181, 182, 183, 184, 185

reported_raw$date <- as.Date(as.numeric(reported_raw$date))

reported_raw$date <- reported_raw$date %m-% years(70) %m-% days(1)

reported_raw<-reported_raw %>%
  mutate(re_categorized_area = case_when(re_categorized_area=="plant A or plant B"~"Both",
                              re_categorized_area=="both"~"Both",
                              re_categorized_area=="field"~"field",
                              re_categorized_area=="logistics"~"logistics",
                              re_categorized_area=="Both"~"Both",
                              re_categorized_area=="plant A"~"plant A",
                              re_categorized_area=="plant B"~"plant B"))

data <- reported_raw %>%
  select(motion_injury,
area_of_injury, ergonomic_risk_factors,
```

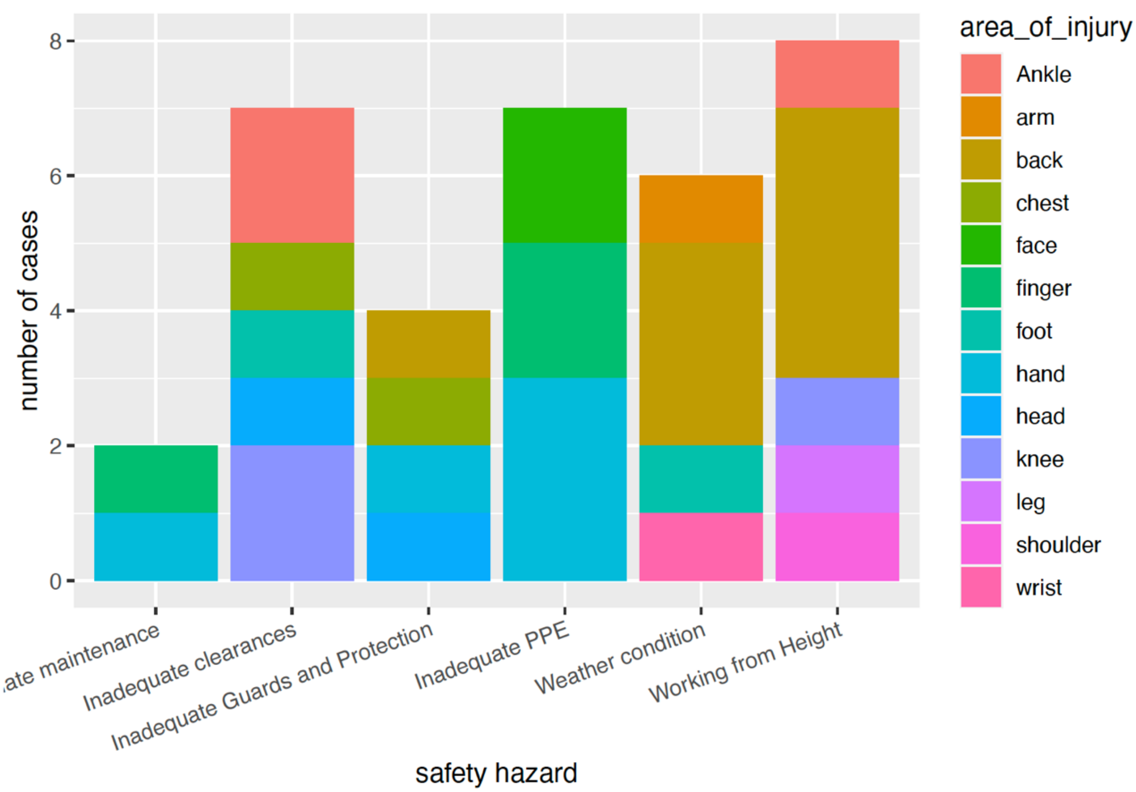
```
                                   motion_injury=="Slipping"~"slipping",
                                   motion_injury=="slippping"~"slipping",
                                   motion_injury=="swinging"~"swinging",
                                   motion_injury=="walking"~"walking"))



data<-data %>%
  mutate(area_of_injury = case_when(area_of_injury=="Abdomen"~"chest",
                                    area_of_injury=="ankle"~"Ankle",
                                    area_of_injury=="arm and legs"~"leg",
                                    area_of_injury=="arm/hand"~"arm",
                                    area_of_injury=="arm/shoudler"~"arm",
                                    area_of_injury=="back/arm"~"back",
                                    area_of_injury=="back/hand"~"back",
                                    area_of_injury=="back/shoulder"~"shoulder",
                                    area_of_injury=="Finger"~"finger",
                                    area_of_injury=="wrist/hand"~"wrist",
                                    area_of_injury=="Lower Back"~"back",
                                    area_of_injury=="back"~"back",
                                    area_of_injury=="Knee"~"knee",
                                    area_of_injury=="wrist/hand"~"wrist",
                                    area_of_injury=="wrist/hand"~"wrist",
                                    area_of_injury=="hand"~"hand",
                                    area_of_injury=="knee"~"knee",
                                    area_of_injury=="finger"~"finger",
                             area_of_injury=="wrist"~"wrist",
                             area_of_injury=="shoulder"~"shoulder",
                             area_of_injury=="Shoulder"~"shoulder",
                             area_of_injury=="arm"~"arm",
                             area_of_injury=="foot"~"foot",
                             area_of_injury=="head"~"head",
                             area_of_injury=="face"~"face",
                             area_of_injury=="finger"~"finger",
                             area_of_injury=="Hand"~"hand",
                             area_of_injury=="Foot"~"foot",
                             area_of_injury=="Ankle"~"Ankle",
                             area_of_injury=="hip"~"hip",
                             area_of_injury=="stomach"~"chest",
                             area_of_injury=="hand/finger"~"hand",
                             area_of_injury=="hand/finger"~"hand",
                             area_of_injury=="Neck"~"head",
                             area_of_injury=="neck"~"head",
                             area_of_injury=="chest"~"chest"))

data<-data %>%
  mutate(ergonomic_risk_factors = case_when(ergonomic_risk_factors=="F/P"~"F/P",
                                            ergonomic_risk_factors=="R/P"~"R/P",
                                                ergonomic_risk_factors=="Forceful Exertions"~"Forceful
                                                ergonomic_risk_factors=="poor posture"~"poor posture",
                                                ergonomic_risk_factors=="repetiton"~"repetiton",
                                                ergonomic_risk_factors=="vibration"~"vibration",
                                                ergonomic_risk_factors=="Vibration"~"vibration"))
```

```
data<-data %>%
  mutate(safety_hazard = case_when(safety_hazard=="inadequate guards"~"Inadequate Guards and Protection
                                    safety_hazard=="inadequate clearances"~"Inadequate clearances",
                                   safety_hazard=="Inadequate clearances"~"Inadequate clearances",
                                   safety_hazard=="inadequate PPE"~"Inadequate PPE",
                                   safety_hazard=="weather condition"~"Weather condition",
                                   safety_hazard=="Weather condition"~"Weather condition",
                                   safety_hazard=="working from height"~"Working from Height",
                                   safety_hazard=="Working from Height"~"Working from Height",
                                   safety_hazard=="inadequate maintenance"~" Inadequate maintenance",
                                   safety_hazard=="inadequate equipment and tool "~"Inadequate Equipme
                                   safety_hazard=="Inadequate training"~"Inadequate training"))

data<-data %>%
  mutate(equipment_source_of_injury = case_when(equipment_source_of_injury=="crane"~"crane",
                                    equipment_source_of_injury=="falling object"~"falling object",
                                   equipment_source_of_injury=="falling objects(panel)"~"falling objec
                                   equipment_source_of_injury=="foreign object"~"foreign object",
                                   equipment_source_of_injury=="Foreign Object"~"foreign object",
                                   equipment_source_of_injury=="heavy objects(truss)"~"heavy object",
                                   equipment_source_of_injury=="machine"~"machine",
                                   equipment_source_of_injury=="machines"~"machine",
                                   equipment_source_of_injury=="nail gun"~"nail gun",
                                   equipment_source_of_injury=="truss"~"heavy object",
                                   equipment_source_of_injury=="Exavator"~"vehicle",
                                   equipment_source_of_injury=="falling objects(wall)"~"falling object
                                   equipment_source_of_injury=="hammer"~"hammer",
                                   equipment_source_of_injury=="ice"~"ice",
                                   equipment_source_of_injury=="metal item"~"metal item",
                                   equipment_source_of_injury=="propane"~"propane",
                                   equipment_source_of_injury=="trusses"~"heavy object",
                                   equipment_source_of_injury=="falling objects(window)"~"falling obje
                              equipment_source_of_injury=="heavy object"~"heavy object",
                            equipment_source_of_injury=="joist"~"heavy object",
                            equipment_source_of_injury=="metal items"~"metal item",
                            equipment_source_of_injury=="roof"~"heavy object",
                            equipment_source_of_injury=="vehicle"~"vehicle",
                              equipment_source_of_injury=="falling object(wall)/trailer"~"falling objec
                            equipment_source_of_injury=="heavy objects(window)"~"heavy object",
                              equipment_source_of_injury=="ladder"~"ladder",
                              equipment_source_of_injury=="mud"~"mud",
                            equipment_source_of_injury=="trailer"~"trailer",
                              equipment_source_of_injury=="wall"~"heavy object",
                              equipment_source_of_injury=="falling objects"~"falling object",
                            equipment_source_of_injury=="Foreign object"~"foreign object",
                            equipment_source_of_injury=="heavy objects"~"heavy object",
                              equipment_source_of_injury=="trailer/falling objects"~"falling object",
                              equipment_source_of_injury=="wrench"~"wrench"))

new_su <- data %>%
  select(motion_injury,
         area_of_injury)
```

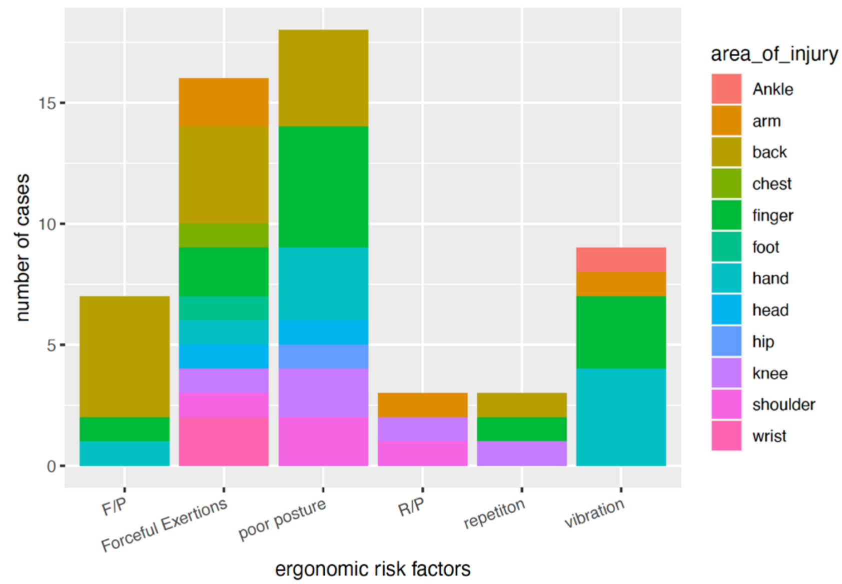
```r
new_su <- na.omit(new_su)

new_su2 <- data %>%
  select(area_of_injury,
         safety_hazard)
new_su2 <- na.omit(new_su2)

new_su3 <- data %>%
  select(area_of_injury,
         ergonomic_risk_factors)
new_su3 <- na.omit(new_su3)

new_su4 <- data %>%
  select(area_of_injury,
         equipment_source_of_injury)
new_su4 <- na.omit(new_su4)

c <- ggplot(new_su, aes(x=motion_injury, fill=area_of_injury)) +
  geom_bar( ) +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1))+
  xlab("motion injury") +
  ylab("number of cases")
c
```



```
b <- ggplot(new_su2, aes(x=safety_hazard, fill=area_of_injury)) +geom_bar( ) +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1)) +xlab("safety
  hazard") +
  ylab("number of cases")
```




```
a <- ggplot(new_su3, aes(x=ergonomic_risk_factors, fill=area_of_injury)) +geom_bar( ) +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1))+
  xlab("ergonomic risk factors") +
  ylab("number of cases")
```




```
d <- ggplot(new_su4, aes(x=equipment_source_of_injury, fill=area_of_injury)) +
  geom_bar( ) +
  theme(axis.text.x = element_text(angle = 20, vjust = 1, hjust=1))+
  xlab("equipment source of injury") +
  ylab("number of cases")
d
```