INTEGRATED DATA ANALYTICS-SIMULATION FRAMEWORK FOR PROACTIVE ASSESSMENT OF SAFETY PERFORMANCE

Estacio Pereira[1], SangUk Han[2], and Simaan AbouRizk[3]

Abstract: Although considerable advances in the proactive control of construction project risks have been reported, identification and assessment of safety-related measures on safety performance remains challenging. This has been attributed to (1) difficulties in data collection; in particular, establishing the number of safety-related measures required to assess their influence on safety performance and (2) difficulties addressing the dynamic nature of projects; in particular, how measures affect safety performance over time. This papers aims to address these issues by implementing a framework that integrates existing departmental data with simulation models to proactively assess and predict safety performance. The framework is composed of three main components. First, safety-related measures available in various departmental databases are identified; second, the relationship between safety performance and measures is analysed and indicators with significant influence are incorporated into the assessment model; and third, a simulation model that reproduces the behaviour of these measures is used to test various scenarios. As evidenced by the results of a case study, the framework proposed here can assist companies with the proactive development of risk-avoidance strategies, thereby improving safety performance.

Keywords: Conceptual safety performance, prediction; policy making; simulation; historical data

1 INTRODUCTION

Over the last several decades, construction worksite accident rates have been considerably reduced. This has been primarily attributed to the implementation of Safety Management Systems (SMS), which function to identify and mitigate causes of worksite accidents. Although functional, SMS have been criticized for their reactive approach: flaws are commonly identified after the occurrence of an accident (Hallowell et al., 2013) and safety performance is usually measured by lagging indicators (Hinze et al, 2013). In response to this criticism, research attention is being directed toward defining safety-related measures that provide early risk warnings, thereby allowing the implementation of proactive, risk-mitigation strategies.

---

[1]   PhD Student, Department of Civil and Environmental Engineering, University of Alberta, 7-203 Donadeo Innovation Centre for Engineering, 9211 116 Street NW, Edmonton, Alberta, Canada, T6G 1H9; estacio@ualberta.ca

[2]   Assistant Professor, Dept. of Civil and Environmental Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Korea, sanguk@hanyang.ac.kr

[3]   Professor, Department of Civil and Environmental Engineering, University of Alberta, 7-203 Donadeo Innovation Centre for Engineering, 9211 116 Street NW, Edmonton, Alberta, Canada, T6G 1H9; abourizk@ualberta.ca

While previous studies have identified several safety-related measures that can proactively control accident risk on construction worksites, these measures are largely associated with safety program performance. However, safety is not disconnected from—and is often affected by—other project components. As such, a broad view of safety-related measures, which considers project elements such as design, schedule, cost, and quality, should be examined. Data collected by departments that are not traditionally associated with safety performance may uncover policies and procedures that enhance SMS. Specifically, analysis of this data could (1) identify the influence of various departmental decisions on safety performance and (2) determine how such data can be used to improve safety management. Furthermore, this data could be used to develop simulation models that allow managers to examine various test scenarios, resulting in a proactive safety management approach that could improve safety performance and reduce accident risk.

To improve the comprehensiveness and proactivity of SMS, this research proposes a framework that assesses and predicts safety performance of construction projects using data from multiple departmental databases.

2 BACKGROUND

To assess safety performance on construction sites, (1) input variables, referred hereafter as safety-related measures, and (2) methods used to assess variable influence on safety performance must be defined.

2.1 Safety Related Measures

SMS can be monitored and controlled by collecting and analysing measures that are associated with construction site accident risk. These measures, referred to as safety-related measures, can monitor the safety level of a system, have predictive value, and can signal interventions to improve safety performance (Hinze et al. 2013).

However, the list of safety-related measures that should be used to control safety performance, in practice, remains inconclusive. In general, influences of measures identified in literature have not been examined from an empirical perspective. These measures are often associated with safety program efficiency (Hallowell *et al.*, 2013), technical aspects of an engineering system, (Han *et al.*, 2014), environment conditions (Lee *et al.*, 2012), worker personnel characteristics (Hinze et al, 2013), and workplace conditions (Goh and Chua 2013).

2.2 Safety predictive models

As reviewed by Esmaeili et al. (2015), several studies employing a variety of methods have attempted to identify relationships between safety-related outcomes (e.g., injury rates, safety climate, perceived risk) and safety-related measures. For instance, Lee et al. (2012) measured project managers' risk perception and used an Analytical Hierarchical Process (AHP) to identify site risk factors and determine their influence on safety risk. Based on elements identified in accident reports, Goh and Chua (2013) used a neural network to assess project safety performance. Esmaeili et al. (2015) recommended a generalized linear model, based on struck-by accident reports, to determine the influence of injury-associated risk factors.

Despite previous efforts to assess safety performance, these proposed methods have difficulty assessing the dynamic behaviour of safety-related measures over time. A framework capable of integrating simulation models with assessment methods, such as

neural networks and AHP, could assist with the proactive evaluation of onsite risk and, consequently, in the development of strategies that improve safety performance.

3 FRAMEWORK

The construction industry has increased its capacity to generate, collect, and store data. This data can provide information from which safety-related measures can be identified and proactively controlled. To analyse the influence of safety-related measures on safety output from existing data, a framework composed of three components, namely a data adaptor, assessment, and simulation model (Figure 1a), is proposed.
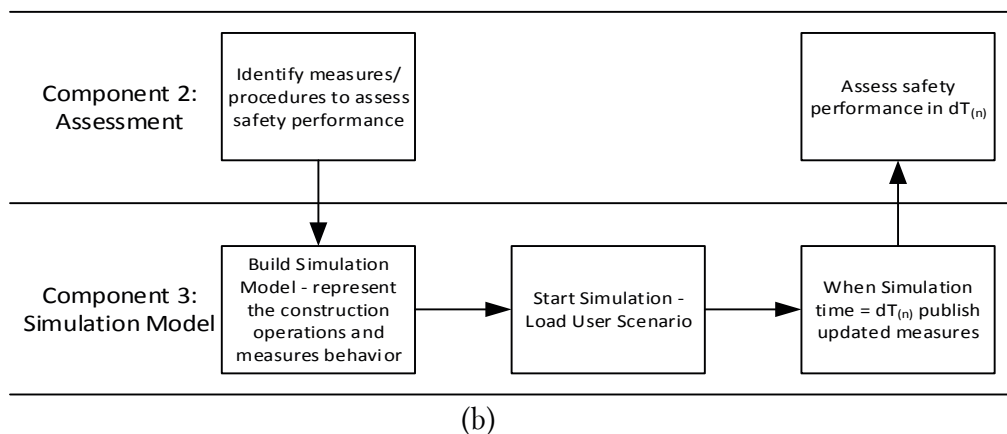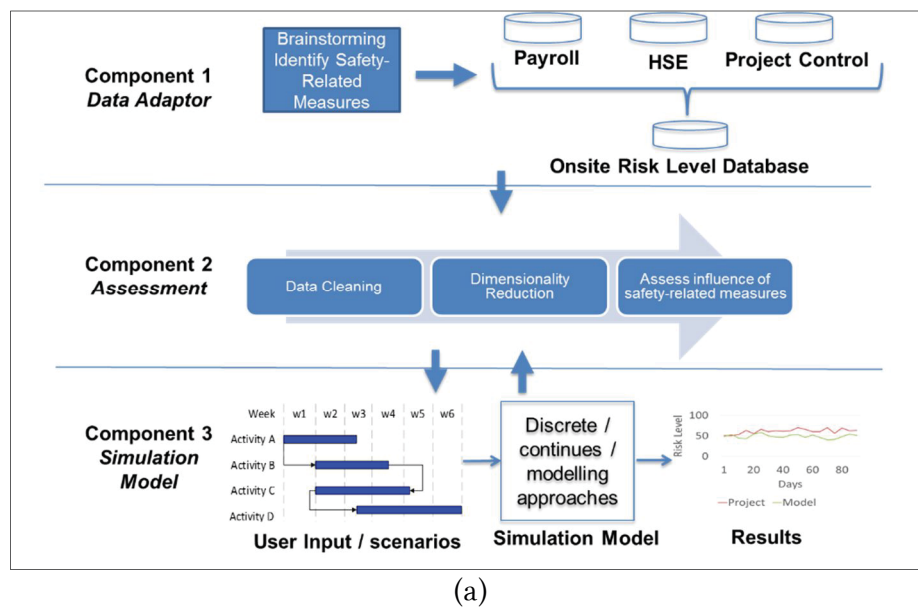

(a)

(b)

Figure 1: a) Framework to assess the safety performance of construction sites; b) Interface between Assessment and Simulation Model components

First, the data adaptor is responsible for collection and integration of safety-related measures from various departmental databases into a centralized, onsite risk level database. The assessment component of the framework defines the influence of each available variable on the output variable and will establish safety performance assessment procedures from these results. Finally, the simulation model is built to mimic the measures identified by the assessment component. Scenarios are then updated by the

user, which updates the value of the measures within the simulation model in the time interval (dT) previously defined. Safety outcome is subsequently assessed based on the procedures defined by component 2 (Figure 1b). A detailed description of each component is provided below.

3.1 Component 1: Data Adaptor

In the proposed framework, measures that may influence safety performance must first be identified. This can be accomplished through brainstorming sessions between project and safety managers. Recently published studies (Hallowell et al., 2013; Han et al., 2014; Esmaeili, Hallowell and Rajagopalan, 2015; Salas and Hallowell, 2016) may guide discussion and identification of measures. For instance, examples of data collected by departments and how they can be considered as safety-related measures is demonstrated in Table 1. Notably, measures can be quantitative or qualitative.

Table 1: Examples of safety-related measures from various departments.

Department	Data Collected	Safety-Related Measures
Payroll	Worker start date on project	New workers rate
		Worker experience on the project
Scheduling	Planned and as-built scheduling	Project hours completed / hours planned
Safety	Inspection dates	Inspection rate
Quality	Hours spent due to rework	Rework hours rate
Cost	Stipend per month	Cost progress / planned cost
Safety	# Behaviour based observation (BBO)	BBO Rate

Data collected by different departments are often stored in various formats within disconnected database systems. In the present framework, data must be collected and integrated within a centralized database. Notably, in addition to identifying and centralizing variables, it is necessary to define metrics. According to Salas and Hallowell (2016), metric definition is important when comparing data with different units. The company should also establish data assessment schedules (e.g., weekly, monthly). The final product of this component is an onsite risk level database that contains all safety-related measures and corresponding safety outputs.

3.2 Component 2: Assessment

In this component, the influence of each safety-related measure is identified and safety performance is assessed. Prior to the determination of measure influence, it is necessary to (1) clean and (2) reduce the number of random variables in the onsite risk level database. In the cleaning process, outliers and missing data are identified and treated accordingly. Following this step, a correlation-based feature selection (CFS) is performed to reduce the dimensionality of the dataset. The primary objective of this test is to decrease the number of safety-related measures while retaining the maximum possible variance. A CFS algorithm was chosen for its ability to deal with a high dimension of safety-related measures and a small number of instances while preserving the original representation of the attributes.

The next step is to define appropriate methods to determine the influence of each safety-related measure on safety output. Methods such as artificial neural networks (ANN), multiple linear regression (MLR), or case-based reasoning (CBR) can be used to establish relationships between input and output variables. Advantages and disadvantages of these methods are summarized in Table 2. The method choice depends on a variety of queries including: is a minimal quantity of data points sufficient; do managers require detailed explanations of the relationships between input and output variables; and is the model continuously updated to improve accuracy?

Table 2: Differences between methods (adapted from Kim, An and Kang, 2004).

Item	MLR	CBR	ANN
Accuracy ranking	3	2	**1**
Clarity explanation between the relationship of input and output variables	Yes	Yes	**No**
Simplicity in updating the model with new data	No	Yes	**No**
Minimal quantity of data points per input required	Yes	No	**Yes**

3.3 Component 3: Simulation Models

In this component, a simulation model is developed to mimic the behaviour of the safety-related measures and—using the method defined in Component 2—predict safety performance. Advantages to using a simulation model to reproduce the behaviour of safety-related measures is its ability to correctly model complex activities and its interactions with required resources; model environmental parameters (such as temperature) from historical data; and include onsite worker availability when calculating productivity rates.

A combined discrete/continuous simulation model is suggested to mimic project behaviour. The discrete portion of the model is responsible for controlling the flow of entities (workers) and its attributes. In the continuous component of the model, time interval is set at one day and the state variable is working hours per day. The continuous component of the model is responsible for regulating the time advance of the entire model. The safety-related measures are updated based on the dT defined by the user.

The user can establish and predict the safety performance of varying scenarios, which will assist the user in their decision-making process. Each scenario should provide information such as project location, project schedule, the relationship between activities, working hours per day, and stochastic distributions to define safety-related measures related to the safety programs.

4 CASE STUDY

A case study from a project in Alberta, Canada is presented to demonstrate the functionality of the proposed framework. The Data Adaptor and Assessment components were based on the findings of Pereira et al. (2017). In this research, a cross-sectional case study was conducted to identify the influence of safety-related measures on the Onsite Risk Level (ORL). Using MLR, the developed model was able to reasonably forecast the safety performance of a project. Equation 1 is used to assess safety performance (Pereira et al. 2017). The results are summarized in Table 3.

$$ORL = (-0.001 * SRM_1) + (0.048 * SRM_2) + (0.275 * SRM_3)$$
$$+(0.081 * SRM_4) + (0.045 * SRM_5) - 8.900 \qquad (1)$$

Where SRM_1 = (BBO) rate; SRM_2 = new workers rate; SRM_3 = workers' age; SRM_4 = near miss rate; SRM_5 = temperature.

Table 3: Summary of data used in the case study (Pereira et al. 2017)

Item	Description
Data points	101 months
Method used	MLR
Databases used	Safety, Financial, Human Resources, and Environment Canada
Output Variable	Onsite Risk Level (ORL)
Safety-related Measures significant	(5) BBO Rate, New workers rate, workers' age, near miss rate, temperature

The primary focus of this case study was to examine the interaction between the simulation model and the MLR. The simulation model aims to reproduce the behaviour of safety-related measures defined by the MLR and construction operations. In the aforementioned case study, the number of working hours required for various activities was summed and distributed over 6 months. Three different scenarios were defined for comparison with actual project safety performance (Figure 2). The production rate to complete the project was modelled using a continuous approach and was equal to the average working hours per month multiplied by the quantity of workers. The new worker's rate varied according to the required number of working hours each month. The distribution type to be used to model the behaviour of the safety-related measures can be defined by (1) a triangular distribution using expert opinion or (2) choosing a distribution that better fits with the data collected by the company (e.g. triangular, beta, etc.) In this case study, the BBO rate was assigned based on a triangular distribution, while the temperature, near miss rate, and workers age remained constant. ORL was assessed on a monthly basis (dT = 1 month).

	Working hours / Month						BBO rate (low, most likely, high)
	Feb	Mar	Apr	May	Jun	Jul	
Actual	90000	115000	125000	135000	143000	142000	(4500, 5200, 6300)
Scenario 1	125000	125000	125000	125000	125000	125000	(3500, 4200, 5000)
Scenario 2	80000	80000	80000	170000	170000	170000	(3000,3200, 4200)
Scenario 3	80000	100000	118000	136000	151000	165000	(4500, 5200, 6300)
Total Working Hours in each scenario: 750000 Average working hours per month/worker = 150 hours / Workers in the field on January = 480							

Figure 2: Scenarios tested in the simulation model

5 RESULTS

Safety-related measures identified by the MLR model were found to influence safety performance over time. The scenarios tested by the simulation model demonstrated that changes in resource allocation and BBO rate affect ORL and, consequently, safety

performance (Figure 3). Overall, the estimated ORL of the actual project was found to be low (< 5). However, if the new worker rate was increased in winter (Scenario 2) or the project was delayed due to a lack of monetary resources (Scenario 3), the ORL peaked at 6 and 8, respectively. Conversely, the actual project ORL was reduced if the BBO rate was increased and the new workers rate was decreased in winter (Scenario 4).
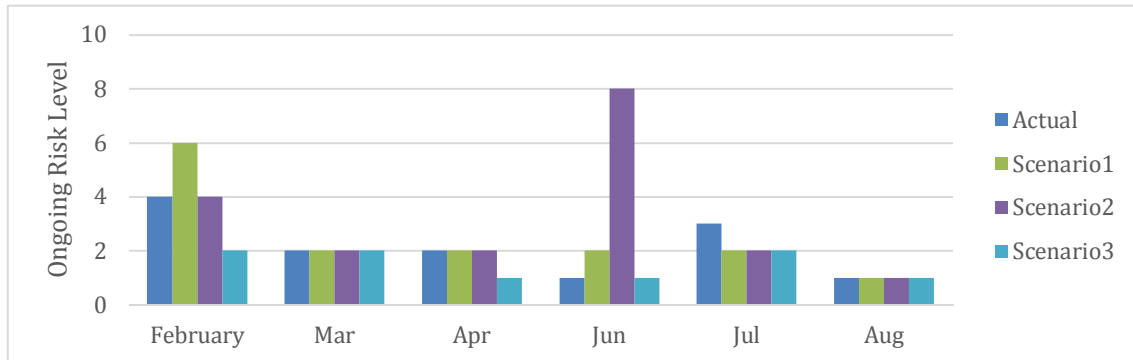

Figure 3: Safety Performance on the different scenarios

The results provided by the framework facilitate the development of risk-avoidance strategies on construction sites. Specifically, interdepartmental data, such as project control and scheduling, can be examined to identify the influence of rework and project delay on safety performance. Furthermore, the proposed framework allows the simulation model to be improved through its consideration of resource constraints, such as equipment and material constraints, and their effect on working hours per month. The model allows companies to proactively tailor safety goals and programs to particular contexts and to reinforce safety procedures relevant to particular project requirements.

Despite the considerable advantages offered by this framework, it should be noted that improvements in model accuracy may be achieved by choosing an alternate assessment technique. As the MLR model is limited to the consideration of dependent and independent variable relationships as linear, other methods such as ANN or CBR may provide more accurate results.

5.1   Simulation model verification and validation

Model validation was performed using a sensitivity analysis. In this procedure, high and low values from each safety-related measure were tested. Results demonstrated that the model output was sensitive to parameters changes. The model behaviour was verified by examining the behaviour of safety-related measures and entities to determine if the model's logic was correct. For this purpose, the simulation tool had a trace window that recorded simulation model events. Information such as time, productivity rate, and updated values of the safety-related measures was verified. The verification results demonstrated that the model implementation was corrected and, therefore, could be used by companies to explore scenarios to proactively control safety performance.

6   CONCLUSIONS

For safety management in construction to become proactive in nature, it is necessary to have a model that reliably predicts the performance and influence of safety-related measures. This study proposes a framework that uses existing data from departmental databases to measure the influence of safety-related measures on safety performance. The proposed framework was found to accurately identify the measures, evaluate their

influence on safety performance, and reproduce their behaviour. The framework was tested through a case study. Here, five measures associated with safety were found to statistically influence safety performance, indicating that resource allocation policies can affect onsite risk level. A simulation model was applied and was found to accurately mimic the behaviour of the safety-related measures. Notably, integration of the MLR and the simulation model allowed safety performance to be assessed over time.

The framework is not intended to predict the number of incidents that will arise during a particular project. Rather, it is developed as a tool to assist companies with the proactive development of safety strategies and to consider of safety performance during project decision-making processes.

7   ACKNOWLEDGMENTS

8   REFERENCES

Esmaeili, B., Hallowell, M. R. and Rajagopalan, B. (2015) 'Attribute-Based Safety Risk Assessment. II: Predicting Safety Outcomes Using Generalized Linear Models', *Journal of Construction Engineering & Management*, 141(8). doi: 10.1061/(ASCE)CO.1943-7862.0000981.

Goh, Y. M. and Chua, D. (2013) 'Neural network analysis of construction safety management systems: a case study in Singapore', *Construction Management and Economics*, 31(5), pp. 460–470. doi: 10.1080/01446193.2013.797095.

Hallowell, M. R., Hinze, J. W., Baud, K. C. and Wehle, A. (2013) 'Proactive construction safety control : measuring , monitoring , and responding to safety leading indicators', *Journal of Construction Engineering and Management*, 139(10), pp. 1–8. doi: 10.1061/(ASCE)CO.1943-7862.0000730.

Han, S., Saba, F., Lee, S., Mohamed, Y. and Pena-Mora, F. (2014) 'Toward an understanding of the impact of production pressure on safety performance in construction operations', *Accident Analysis and Prevention*, 68(7), pp. 106–116. doi: 10.1016/j.aap.2013.10.007.

Hinze, J., Thurman, S. and Wehle, A. (2013) 'Leading indicators of construction safety performance', *Safety Science*, 51(1), pp. 23–28. doi: 10.1016/j.ssci.2012.05.016.

Kim, G.-H., An, S.-H. and Kang, K.-I. (2004) 'Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning', *Building and Environment*, 39, pp. 1235–1242. doi: 10.1016/j.buildenv.2004.02.013.

Lee, H., Kim, H., Park, M., Teo, E. A. L. and Lee, K. (2012) 'Construction risk assessment using site influence factors', Journal of computing in Civil Engineering, 26(3), pp. 319–330. doi: 10.1061/(ASCE)CP.1943-5487.0000146.

Salas, R. and Hallowell, M. (2016) 'Predictive Validity of Safety Leading Indicators: Empirical Assessment in the Oil and Gas Sector', *Journal of Construction Engineering & Management*. doi: 10.1061/(ASCE)CO.1943-7862.0001167.