INVESTIGATIVE MODEL OF RAIL ACCIDENT AND INCIDENT CAUSES USING STATISTICAL MODELLING APPROACH

SHAMSULLARIFIN BIN ISMAIL

A project report submitted in partial
Fulfilment of the requirement for the award of the
Master of Science in Railway Engineering

Faculty of Engineering Technology
Universiti Tun Hussein Onn Malaysia

July 2016

ACKNOWLEDGEMENTS

Alhamdulillah, first and above all, I praise to Allah S.W.T the Almighty for providing me this opportunity and granting me the ability to proceed successfully.

Here, I would like to express my sincere gratitude to my supervisor Dr. Joewono Prasetijo for the continuous support towards my master study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master study.

Besides my supervisor, I would like to thank the rest of my thesis committee, Prof. Madya Dr Khalid, Prof. Madya Dr Zainal Abidin, and Dr. Basil for their insightful comments and encouragement, but also for the hard question which incanted me to widen my research from various perspectives.

I would also like to extend my gratitude towards all master students under Dr Joewono especially to Wan Zahidah, Mohd Farhan, Natasha and Noraini. I greatly appreciates their kindness to advise me and share their opinion during the period where I am but lost in finding my destination to complete my research. Thanks also to all the members of master science in railways engineering including all staff member.

Here also I would like to thanks to industrial representative from KTMB and SPAD. They continuously showed their interest and willing to entertain me and accept student doing research together with them. Event collaboration was not met due to confidential issue, but they always helped me with useful advice and idea. Last but not the least, I would like to thank my family, my parents and to all my brothers and sister for supporting me spiritually throughout the span of this thesis and my life in general.

ABSTRACT

Nowadays, railway transportation becomes a popular choice among commuter as transportation mode to travel from one place to another. Thus, it makes the industry grows faster especially at urban area. The complexity of rail network required high level of safety features to prevent any interruption. For that purpose, this thesis will show a proper procedure on how the prediction model of accident need to be conducted using regression model. From the root cause analysis, the most contributory factor to influence the accident can be identified. "Ishikawa diagram" is a popular tool to identify problem occurring from the root where it begins. Process of identifying required bundles of accident and incident investigation report at least for 5 years and this thesis used data starting from 1999 to 2014. It was taken from several sources on Australian Railways website. Analysis from Ishikawa shows there are ten main factors involved to influences an accident. Those factors are "train driver mistake", "other's human mistake", "weather influence", "track problem", "train problem", "signaling error", "maintenance error", "communication error", "procedure error", and "others". Each factor with positive correlation coefficient value to the type of accident and incident were taken as parameter. Then, before completing the prediction model formula, some of hypothesis needs to be tested to know which model among regression model is suitable and give a better prediction result. Dispersion test is a test to calculate dispersion value to know either data is under dispersion for value less than 1 (Poisson model is appropriate) or over dispersion for value more than 1 (Negative binomial is appropriate). Then, Vuong test is applied to measure which model has a better result between those two models. From the hypothesis, this thesis shows that Zero-inflated model is the most fitted model to predict accident and incident cases of collision, derailment and SPAD. In some country, they may have different system of rail and geography, thus it should have different possibilities to influence accident and incident. However, this method and procedure are available to use for them to identify and predict the most influencing factor that contributes to the accident occurrences.

CONTENTS

	TITLE		**i**
	DECLARATION		**ii**
	ACKNOWLEDGEMENT		**iii**
	ABSTRACT		**iv**
	CONTENTS		**v**
	LIST OF TABLES		**vii**
	LIST OF FIGURES		**viii**
	ABBREVIATION		**ix**
	LIST OF APPENDICES		**x**
CHAPTER 1	**INTRODUCTION**		**1**
	1.1	Research Background	1
	1.2	Problem Statement	3
	1.3	Objective	3
	1.4	Scope of study	4
CHAPTER 2	**LITERATURE REVIEW**		**5**
	2.1	Rail accident & incident terminologies	5
	2.2	Accident identification analysis	6
	2.3	Prediction formula as modelling approach	10
		2.3.1 Peabody and Dimmick formula	10
		2.3.2 The New Hampshire Index formula	11
		2.3.3 The NCHRP hazard index formula	13
		2.3.4 USDOT accident prediction model	14
	2.4	Regression technique	17
		2.4.1 Poisson regression model	17
		2.4.2 Negative binomial regression model	18
		2.4.3 Zero-truncated model	19
		2.4.4 Parameter use in modelling approach	20

2.5 Summary 21

CHAPTER 3 RESEARCH METHODOLOGY 23

3.1 Introduction 23

3.2 Collecting data 26

3.3 Extracting data to Microsoft Excel 27

3.4 Root cause analysis using Ishikawa diagram 30

3.5 Develop prediction model 31

3.5.1 Correlation coefficient value 31

3.5.2 Regression model 32

3.5.3 Dispersion test 33

3.5.4 Vuong test 34

3.5.5 Estimated coefficient value 34

CHAPTER 4 CASE STUDY / DATA ANALYSIS 35

4.1 Introduction 35

4.2 Ishikawa diagram 35

4.3 Correlation coefficient value 40

4.4 Poisson model 41

4.5 Dispersion test 43

4.6 Zero-inflated Poisson model 44

4.7 Vuong test 46

4.8 Summary 46

CHAPTER 5 CONCLUSION AND RECOMMENDATION 48

5.1 Conclusion 48

5.2 Limitation 51

5.3 Suggestion and recommendation 51

REFERENCES 52

APPENDICES 55

LIST OF TABLE

2.1	List of major transit accident types and possible causes	6
2.2	Vehicles per day factor (NCHRP Hazard Index)	13
2.3	Existing devices factor (NCHRP Hazard Index)	14
2.4	Variables for Equation 14 (USDOT Prediction Model)	15
2.5	Rural highway type Values (USDOT Prediction Model)	15
2.6	Urban highway type Values (USDOT Prediction Model)	16
3.1	Number of transport involve in accident and incident	27
3.2	Accident and incident of rail categories by it element	28
4.1	Correlation coefficient value for all variable	40
4.2	Dispersion test	43
4.3	Vuong test result	46

LIST OF FIGURE

2.1	Basic diagram of Ishikawa diagram or fishbone	8
2.2	Two-stage analysis for the identification of hazards in the rail rapid transit	9
3.1	Flow of research	24
3.2	Flowchart of extracting data	25
3.3	Percentage of accident cases by categories	27
3.4	Accident and Incident of rail by type and category	28
3.5	Accident and incident of rail by categories and element influence	29
3.6	Sample of Ishikawa diagram	31
4.1	Ishikawa diagram of collision	36
4.2	Ishikawa diagram of derailment	37
4.3	Ishikawa diagram of SPAD	38
4.4	Number of accident cases and it percentage for each contributory factor	39
4.5	Correlation of factor influence to the accident and incident category	41

ABBREVIATION

AJTP	-	ASEAN-Japan Transport Partnership
ATS	-	Automatic Train Supervision
ATP	-	Automatic Train Protection
ATSB	-	The Australian Transport Safety Bureau
FHWA	-	The Federal Highway Administration
FMEA	-	Failure Modes and Effects Analysis
HSIP	-	Highway Safety Improvement Program
NCHRP	-	National Cooperative Highway Research Program
OTSI	-	The Office of Transport Safety Investigations
RCA	-	Root cause analysis
SPAD	-	Signal Pass at Danger
USDOT	-	United State Department of Transport
ZIP	-	Zero-inflated Poisson
NB	-	Negative Binomial
ZTNB	-	Zero-truncated Negative Binomial

LIST OF APPENDICES

APPENDIX	TITLE	PAGES
A1-A3	Summary of Accident Prediction/Hazard Index Formula	55
B	Symbol of formula	58
C1-C6	List of Accident and Incident Investigation Report from The Australian Transport Safety Bureau (Atsb)	59
D	List of Accident and Incident Investigation Report from Department of Transport and Main Road, Queensland Government	65
E1-E3	List of Accident and Incident Investigation Report from Department of Economic Development, Jobs, Transport and Resources	66
F1-F3	List of Accident and Incident Investigation Report from Office of Transport Safety Investigations (OTSI)	69
G1-G28	Summary of Data Collection	72
H1-H8	Simulating Prediction Model Equation Using R-Studio	100

CHAPTER 1

INTRODUCTION

1.1 Project background

Rail transportation is one of the largest transportation modes in the world. This transportation was very popular among developed country. It also has multi-purpose function. Early of their invention, rail transportation was use to bring coal, ore and others industrial goods to the warehouse and port. As time advanced, these modes have evolved from hand power to the horse power, then improve to steam engine, diesel engine, and nowadays are electrified system. Rail system is not only to bring goods, but also as transportation mode to move people to their destiny. This mode is also the best solution for public transport especially at urban area because of their advantages. By commuting with rail service, consumer is able to manage their time journey, reducing transportation cost, able to avoids traffic jam, save time, reduce the environment pollution, and many more. Besides that, this rail system also can move in large number of passengers per trip, especially in big city and urban areas. All the advantages and higher demands make a lot of effort by researcher and among rail manufacturer such as Bombardier (Canada), Alstom (France), and Siemens (Germany) to compete each other's to advancing their technologies. Thus, governments spend billion dollar of money to invest in these technologies. Urban light rail systems and subways are expanding in many regions of the world, and there is growing investment in intercity high-speed rail lines[1].

Rapidity of rail system makes it more complicated and complex. A lot of safety devices were introduced to improve previous system. There is not only focus on signaling system and equipment such as interlocking and block signal, but also to the boogie system, infrastructure and many more by the manufacturer. All the element of rail must be in good condition and well function to avoid any possibilities of accident occur. In that case, risk assessment is important in initial planning to improve rail network in terms of safety for the future. Thus, periodical statistical data is required to monitor performance and efficiency of rail services. Rail accident and incident is a major factor that will cause interruption to the rail operation. For that, more research that focused on this issue is needed. Such as in the United States, the Federal Highway Administration (FHWA) periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement at rail road crossing. They require each state to develop and implement a highway safety-improvement program (HSIP), which consists of three components. Those are planning, implementation, and evaluation [2].

However, preventing action is not only at rail road areas, periodical surveying or monitoring to the others section also important to prevent any accident occur such as rail road structure, geometry and surrounding condition, maintaining program, working environment and so on. All of this must be planned properly to ensure entire system in good condition without any disturbances. Especially at rural area, lower slopes of hill, bridges, tunnel area, and also at the middle of jungle. Thus, continues effort to increase safety level at all aspect is required to ensure zero cases in the future. Prevention action is not to prevent human soul and injury only, but this program also important to prevent any loses to the facilities such as train and track.

By doing deep analysis, root cause leading to the accident and incident events can be detected. Furthermore, effectiveness of counter measure had been taken before it can be evaluated. Thus, further effort can be done properly to improve safety performance and eliminate error occurrence from beginning. By using mathematical modeling approach, correlation and coefficient value of each factor can be calculated. Prediction model is used to improve monitoring system of railway network to overcome any danger and possibilities of train accident in the future. Further information of prediction model use in railway is discussed on chapter 2. Other than that, statistical data will help management to maintain and monitor the safety performance of the rail track network. Furthermore, special program to motivate employee can be arranged to ensure they always in good condition either physical or mental during on job. Thus, Incident cases cause by human can be neglected. In terms of tangible benefit, they can prevent their loss of profit towards restructuring the disaster area caused by an accident.

1.2 Problem statement

Risk assessment related to the rail industries had been done since 1941. Until now, the assessment only focused on collision at rail road crossing with warning devices and track features as a subject or parameter. However, it is less researched at the other area. Accident or incident can be happen at anywhere with various conditions. Accident and incident could be happening, which might be led by device failure, procedural mistake, or any interruption. Furthermore, others accident or incident categories such as derailment and signal pass at danger also can give possibilities to the loss of human soul or injury. Since every country has a different geographic and operating system of rail in their networks, prediction model could be difficult to be fixed for their independent variable or factor influence. Because of that, there is no proper investigative procedure to predict the rail accident and incident.

Thus, this paper will focus on method how to clarify the most influence factor will affect to the accident or incident event and become independent variable for prediction model. It is combination between root cause analysis to find the contributory factor and prediction model as a predictor. Thus, priority of each factor can be found. Since, there are a lot of model for regression, this procedure also to identify which model is most suitable to use after go through certain steps and hypothesis test. Then, this method can be standardizing in rail industry everywhere.

1.3 Objective

1. To develop investigative procedure to predict factor influence of railway accident.
2. To identify contributory factor that influencing to the rail accident using Ishikawa diagram through case study of railway accident investigation report.
3. To formulate the prediction model of railway accident and incident using regression formula.

1.4 Scope of study

This thesis focuses on case study of railway accident investigation report had been officially announce and publish for public reviews. The purpose of case study is to investigate the factor that causing those accident occurring without blaming any parties. The case study report use in this thesis was started from 2009 to 2014. There were several official government webpages in Australia was use as reference to catch the data. This selection was made because in some country, this information is prohibited and confidential. Then, the data extracted will be categories to put in Ishikawa diagram to find the root cause. For that, concept of Ishikawa diagram must be understood. The main factor taken from Ishikawa and statistical data from case study was use to predict the most influence factor causes by using regression formula. To simulate regression formula, R-studio software was used to find the correlation value and coefficient value. Before that, it must follow certain process and hypotheses test. Further detail methodology was discussed on Chapter 3.

CHAPTER 2

LITERITURE REVIEW

2.1 Rail accident and incident terminologies

Referring to the “Federal Railroad Administrative Office of Safety Analysis (FRA)”, they use term of "Accident/Incident" to describe the entire list of reportable events. These include collisions, derailments, and other events involving the operation of on-track equipment and causing reportable damage above an established threshold, impacts between railroad on-track equipment and highway users at crossings, and all other incidents or exposures that cause a fatality or injury to any person, or an occupational illness to a railroad employee. These terminologies are divided into three major groups for reporting purposes[3].

1. Train accidents

   A safety-related event involving on-track rail equipment (both standing and moving), causing monetary damage to the rail equipment and track above a prescribed amount
2. Highway-rail grade crossing incidents.

   Any impact between a rail and highway user (both motor vehicles and other users of the crossing as a designated crossing site, including walkways, sidewalks, etc., associated with the crossing
3. Other incidents.

   Any death, injury, or occupational illness of a railroad employee that is not the result of a "train accident' or "highway-rail incident".

However, “New Zealand Railways Act 2005” define accident as an occurrence associated with the operation of a rail vehicle or the use of railway infrastructure or railway premises that causes such as the death or serious injury to individuals, or significant damage to property. While Incident means an occurrence other than an accident that is associated with the operation of a rail vehicle or the use of railway infrastructure or railway premises that placed or could have placed such as a person at risk of death or serious injury or property at risk of serious damage[4].

2.2 Accident identification analysis

Managing risk assessment is one of the ways to improve safety of railway by identifying or predicting the derailment. Zarembski [5] in their research had focused on risk assessment of broken rail, rail buckling and vehicle track geometry. By implementing all risk assessment model to this element in Europe and US country, it showed a positive impact to the reduction of failure/defect about 28% - 33% for broken rail and 30% - 50% reduction of rail buckling[5]. Continues effort of this activities to identify and prediction of hazardous is needed to ensure this positive impact will increase and minimize the possibilities of an accident from happening again.

Several researcher in their study [6][7][8] has mention for the possible causes of transit accidents usually happen in previous event has summarize in Table 2.1.

Table 2.1: List of major transit accident types and possible causes

Type	Possible causes
Fire	failure of electrical equipment, brakes operation, storage of inflammable material, derailing, the strike of lightening
Flood	Poor design or failure of the drainage system
Collisions	1. Violation of stop signs by the driver (intrusion) 2. Failure of the ATC system 3. Signal errors (warning devices) 4. Breakdown or mishandling of the split switch 5. Train speed 6. Number of train operation and vehicle (ADT)
Derailment	1. Incomplete release of hand brake 2. Inadequate geometric design 3. Excessive rocking of the trains 4. Speeding at the turn 5. Damage of the bearing and wheel 6. Main operation 7. Train handling 8. Track geometry 9. Buckled track

	10. Main line break operation
Door Accident	Jammed by the door or the uncontrolled opening of the door
Breakdowns of power supply system	Failure of the power supply or emergency power supply facilities, overload, mishandling, or struck by lightening
Intrusion	Animals or people could be found intruding the track on the level ground or the station
Gap fall	Inadequate design of the platform or station
Scraped by the train	Mindless passenger or lack of proper signs on the platform Natural
Natural Disaster	Earthquakes, lightening, storms, or heat waves
Others	Criminal Acts, suicide, crowding, etc.

This accident and incident possibly can be categories in three main elements or categories and it would be considered as contribution factor to the hazards of railway accident [6]. Those are:

1. Human: operators and passengers
2. System: material, equipment, tools, and safety facility
3. Environment: temperature, humidity, ventilation, lights, and noises (natural environment and artificial environment) in the working place.

Inter relation between entire elements also make the possibilities become huge to influencing the accident. For that, deep analysis when the accidents happen is required to find out the root cause.

According to Vorley [9], Root Cause Analysis (RCA) is a method that is used to address a problem or non-conformance, in order to get to the “root cause” of the problem. It is used so we can correct or eliminate the cause, and prevent the problem from recurring. While Mahto & kumar [10] mention Root Cause Analysis (RCA) is the process of identifying causal factors using a structured approach with techniques designed to provide a focus for identifying and resolving problems. The most popular tools usually used in RCA are Ishikawa diagram or fishbone charts to measure cause-effect relationship of how these factors lead to hazards. Below Figure 2.1 show a basic diagram for Ishikawa diagram tool[9].

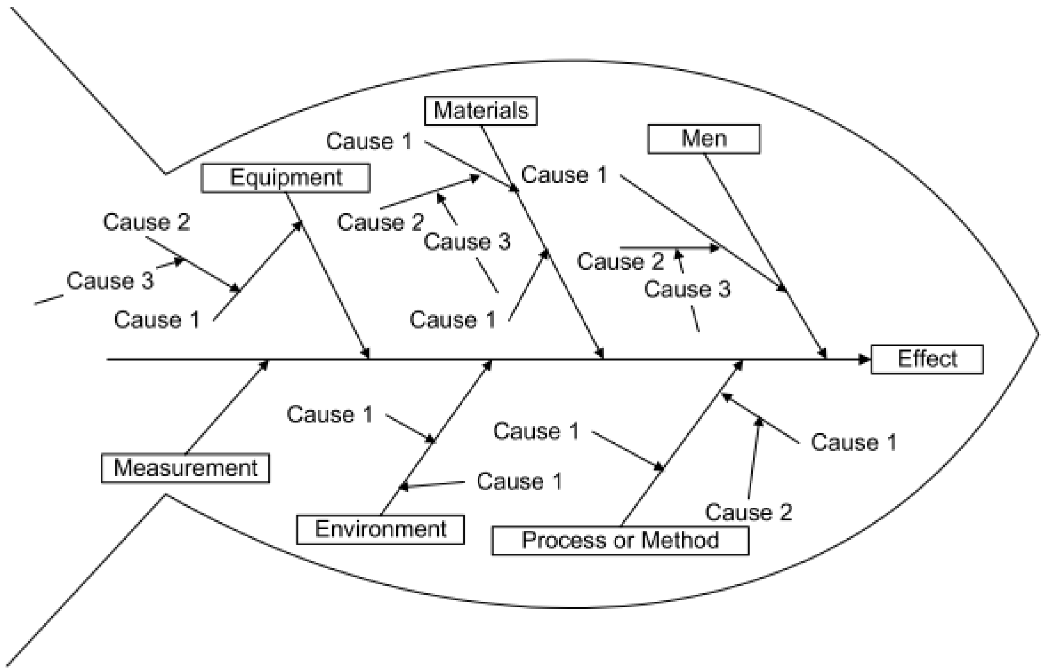

Figure 2.1: Basic diagram of Ishikawa diagram or fish bone[9]

The others tool is Failure Modes and Effects Analysis (FMEA), Failure modes and effects analysis (FMEA) is an evaluation technique used to identify and eliminate known and/or potential failures, problems, and errors from a system, design, process, and/or service before they actually occur. FMEA is to prevent errors by attempting to identify all the ways a process could fail, estimate the probability and consequences of each failure, and then take action to prevent the potential failures from occurring. According to the Hugest [10], Root cause analysis (RCA) used extensively in engineering and similar to critical incident technique is a formalized investigation and problem-solving approach focused on identifying and understanding the underlying causes of an event as well as potential events that were intercepted. RCA is a technique used to identify trends and assess risk that can be used whenever human error is suspected with the understanding that system, rather than individual factors, are likely the root cause of most problems. A similar procedure is critical incident technique, where after an event occurs, information is collected on the causes and actions that led to the event.

According to Wang [6] in their study found that there are two stages to identify the accident that show in Figure 2.2. At the first stages, all the element factors would cause the system unsafe conditions, while they might not affect the normal operation immediately. Under such a situation, the specific unsafe conditions or behaviors can trace back and identify the inherent hazards. In this case, most of the hazards can be eliminated under a perfect management system, for example: an accurate routine check system, the alertness of the operators, a good auditing system, and a strict quality and safety control system. Necessary defenses or barriers are the last shift use to keep the unsafe situations from adverse consequences. If the unsafe situations keep on remaining in the system, the barriers themselves are identified as the source of direct hazards in the second stage. When the potential hazards are not eliminated in the first stage, the automatic train operation control and monitoring system will come into play. The Automatic Train Supervision (ATS) and Automatic Train Protection (ATP) should detect any abnormalities in the system and send the messages to the control center for correction or to the driver to pull on the emergency brake. Once the unsafe conditions keep on exposing, the hazards in the second stage will be identified by the factors that the automatic train control and monitoring system fail to respond with safety measures and stop the exposure of hazardous elements.

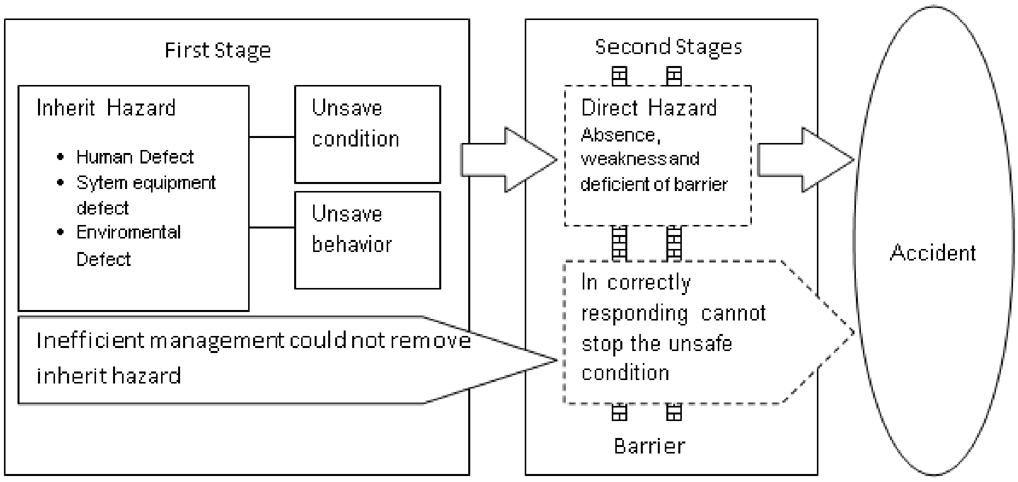

Figure 2.2: Two-stage analysis for the identification of hazards in the rail rapid transit [6]

2.3 Prediction formula as modeling approach

Since early 1900's, Rail network has growing up in number and territory to connecting entire nation especially in Europe, United States, India, and others country as a main transportation to carrying people and things. Interruption to the rail track which has own right of way will cause disaster. In United States, the exposure to potential collisions between trains and motor vehicles at some 224,000 railroad-highway grade crossings has created a serious problem with regard to the convenience and safety of highway travel [11]. For that, many activities had been done to improve safety in order reducing an accident especially in number of death. Thus, as an effort for that, some devices has installed to the railroad-highway area as a signal or warning to make people alert the situation where the train will be crossing that area. It is not only for human, but it also for the system that controlling the crossing area as a system protection such as flash light and short arm gates with flash light. Since 1941, Peabody and Dimmick start to study the potential of hazard as an approach to reducing the number of accident and until now currently this effort has continues to be study.

2.3.1 Peabody and Dimmick formula

L. E. Peabody and T. P. Dimmick, in a 1941 study performed by the Division of Transport, Public Roads Administration, collected data on 3,563 crossings in 23 states for a five-year study period. The prediction of accident frequency is useful both in the determination of crossing warrants and for the economic justification of crossing protection. The prediction equation proposed by Peabody and Dimmick is as follows:[12]

$$P = \frac{1}{N}\sum\frac{(H)(T)}{100A} = \frac{1}{100N}\sum\frac{(H)(T)}{A} \tag{2.1}$$

Where;

P = the protection coefficient for a type of protection

N = the number of crossings in a type group

H = the daily highway traffic volume at each crossing

T = daily train traffic volume at each crossing,

A = number of accidents.

A correlation analysis was used to develop the following equation:[11]

$$I = 1.28\frac{(H^{0.170})(T^{0.151})}{P^{0.171}} + K \qquad (2.2)$$

2.3.2 The New Hampshire index formula

After the Peabody-Dimmick formula was published, a number of cities and states developed their own hazard index formulas and methods for use in ranking railroad crossings for safety improvements (Referring to the APPENDIX A1-A3 & B for more detail) [13]. The study [14] has mention that the New Hampshire formula to be the optimum formula to use as a start towards developing a railroad crossing safety improvement program. New Hampshire formula is the most straightforward and uses three readily available inputs. The New Hampshire Index is a very simple hazard index calculation that can give a high level ranking to determine the need and relative priority of railroad crossings for safety improvements. Based on this formula, railroad crossings with higher exposure factors and/or passive warning devices will rank as a higher priority for safety improvements than will railroad crossings with lower exposure factors and/or more active levels of warning devices. The New Hampshire Index formula is:

$$HI = (V)(T)(P_f) \qquad (2.3)$$

Where;

HI = hazard index

V = AADT volume

T = average daily train traffic volume

P_f = protection factor (0.1 for gates, 0.6 for flashing lights, and 1.0 for signs only)

Austin[15] in their report has mentioned that several states have developed their own variations of the New Hampshire Index with utilizing the basic equation with modifications to allow incorporation of other accident causative factors.

Variations of the New Hampshire Index follow:

Variation 1: $$HI = (V)(2T_f)(T_s)\frac{(SD+AN+NTR)}{4} \qquad (2.4)$$

Variation 2: $$HI = \frac{(V)(T)(A_5)}{P_f} \qquad (2.5)$$

Variation 3: $$HI = (V)(T)\frac{(TT)+TTR+SD+AN+AL+L+G+VSD+W+LI)}{100} \quad (2.6)$$

Variation 4: $$HI = \frac{(P_f)(V_f)(T)(TS)(NTR)}{160} + (70A_a)^2 + 1.2(SD) \quad (2.7)$$

$$A_a = \left(V + \frac{SBP}{1.2}\right)(HM) \quad (2.8)$$

Variation 5: $$HI = 0.1(P_f)(A_f)(T_1) + (AN)(NTR)(S)(0.5L) + TS((FC)(P) + \left(\frac{(V)(T)}{10000}\right) + SB) \quad (2.9)$$

Variation 6: $$HI = \frac{(V_f)(P_f)(T)}{(TR+TN+T_f+HS+G+SD+AN)} \quad (2.10)$$

Variation 7: $$HI = 0.1(V)(T) + 0.1(HS)(TS) + (SD)(AN)(TR)(NTR)(AL) + ({A_a}^2 + 1)(RF)(LP)(P_f) + (SB)(SBP) + 10(HM \quad (2.11)$$

Variation 8: $$HI = \frac{(T)\sqrt{(V)}}{P_f} \quad (2.12)$$

A_s = number of accidents in five years
A_a= number of accidents per year
A_f = accident factor
AL = highway alignment factor
AN = approach angle factor
FC = functional class factor
G = approach grades factor
HI = hazard index
S = surface type factor
SB = number of school buses
SBP=Uumber of school bus passengers
SD = sight distance factor
T = average number of trains per day
T_f = number of fast trains
TN = number of night trains factor
TR = number and type of tracks factor
HM = hazardous material vehicles factor
TS = train speeds factor
HS = highway speed factor
L = number of lanes factor
LI = local interference factor
LP = local priority factor
NTR = number of tracks factor
P = population factor
P_f = protection factor
T_s = number of slow trains
TT = type of train movements factor
TTR = type of tracks factor
V = annual average daily traffic
V_f = annual average daily traffic factor
VSD = vertical sight distance factor
W = crossing width factor
RF = rideability factor

The dissimilarity between the New Hampshire Index model variations raises concerns over its validity. While most of the discrepancies can be attributed to state preferences, concern is raised due to the lack of consistency. Depending on the variation chosen, prediction values vary considerably.

2.3.3 The NCHRP hazard index formula

The National Cooperative Highway Research Program (NCHRP) Hazard Index, documented in NCHRP Report 50, was published in 1964 in a joint effort between the American Association of State Highway Officials (AASHO now AASHTO) and the Association of American Railroads (AAR) in response to the disproportionately high number of accidents occurring at highway-rail crossings. The NCHRP Hazard Index used accident data that spanned five years and was collected by the Interstate Commerce Commission, state agencies and others[15].

The NCHRP Hazard Index closely resembles the basic formula of the New Hampshire Index described above:

$$EA = (A)(B)(CTD) \tag{2.13}$$

Where;

EA = expected crash frequency

A = vehicles per day factor (refer Table 2.2)

B = protection factor indicative of warning devices present (refer Table 2.3)

CTD = current trains per day

Table 2.2: Vehicles per Day Factor (NCHRP Hazard Index)[15]

Vehicle per day	A	Vehicle per day	A	Vehicle per day	A
250	0.000347	6000	0.007720	16000	0.019549
500	0.000694	7000	0.009005	18000	0.021736
1000	0.001377	8000	0.010278	20000	0.023877
2000	0.002627	9000	0.011435	25000	0.029051
3000	0.003981	10000	0.012674	30000	0.034757
4000	0.005208	12000	0.015012		
5000	0.006516	14000	0.017315		

Table 2.3: Existing Devices Factor (NCHRP Hazard Index) [15]

Existing Devices		B
A	Cross bucks, highway volume less than 500 per day	3.89
B	Cross bucks, urban	3.06
C	Cross bucks, rural	3.08
D	Stop signs, highways volume less than 500 per day	4.51
E	Stop signs	1.15
F	Wigwags	0.61
G	Flashing lights, urban	0.23
H	Flashing lights, rural	0.93
I	Gates, urban	0.08
J	Gates, rural	0.19

The NCHKP Hazard Index is concise and easy to use. Unfortunately, this is both its virtue and its vice. There are only three variables to calculate which makes it easy to use, but this limits its descriptive capabilities.

2.3.4 USDOT Accident Prediction Formula

The USDOT Accident Prediction Formula, developed in the early 1980's, is most widely used. This complex and comprehensive formula comprises three primary equations:[15]

$$a = (K)(EI)(DT)(MS)(MT)(HP)(HL)(HT) \qquad (2.14)$$

The factors in Equation 14 each represent characteristics of crossings in the Rail-Highway Crossing Inventory (see Tables 2.4, 2.5 and 2.6). These factors were found to be statistically significant, using nonlinear multiple regression, in the prediction of accidents at highway-rail crossings. Factors such as sight distance are unavailable in the Rail-Highway Crossing Inventory. Using Table 2.4, the value calculated represents the factor's influence in the prediction of accidents at highway-rail crossings where:

c = number of highway vehicles per day
t = number of trains per day
mt = number of main tracks

d = number of through trains per day during daylight
hp = highway payed (yes = I and no = 2.0)
ms = maximum timetable speed in mph
hi = number of highway lanes
ht = highway type factor (see Tables 2.5 and 2.6 below)

Table 2.4: Variables for Equation 2.14 (USDOT Accident Prediction Model)[15]

Variable	Description	Coefficient or Relationship Variable		
		Passive Control	Flashing Lights	Gates
K	Formula Constant	0.002268	0.003646	0.001088
EI	Exposure Index Factor	$(\frac{ct+0.2}{0.2})^{0.03334}$	$(\frac{ct+0.2}{0.2})^{0.2953}$	$(\frac{ct+0.2}{0.2})^{0.3116}$
DT	Day Through Trains Factor	$(\frac{d+0.2}{0.2})^{0.1336}$	$(\frac{d+0.2}{0.2})^{0.2953}$	1.0
MS	Maximum Speed Factor	$e^{0.0077ms}$	1.0	1.0
MT	Highway Paved Factor	$e^{0.2094mt}$	$e^{0.1088mt}$	$e^{0.2912mt}$
HP	Highway Paved Factor	$e^{-0.6160(hp-1)}$	1.0	1.0
HL	Highway Lanes Factor	1.0	$e^{0.1380(hl-1)}$	$e^{0.1036(hl-1)}$
HT	Highway Type Factor	$e^{-0.1000(ht-1)}$	1.0	1.0

Table 2.5: Rural Highway Type Values (USDOT Accident Prediction Model)[15]

Highway Type: Rural	Highway Type Factor (ht)
Interstate	1
Other principal arterial	2
Minor arterial	3
Major collector	4
Minor collector	5
Local	6

Table 2.6: Urban Highway Type Values (USDOT Accident Prediction Model)[15]

Highway Type: Urban	Highway Type Factor (ht)

Interstate	1
Other freeway/expressway	2
Other principal arterial	3
Minor arterial	4
Collector	5
Local	6

$$B = \frac{T_o}{T_o+T}(a) = \frac{T_o}{T_o+T}\left(\frac{N}{T}\right) \quad (2.15)$$

Equation 2.15 adjusts the accident prediction value a, from Equation 2.14 to reflect the actual accident history at the crossing. The variable, N, is the number of observed accidents in T years at the crossing, and T_0 is the formula weighting factor defined as:

$$T_o = \frac{1.0}{(0.05+a)} \quad (2.16)$$

$$A = (k)(B) \quad (2.17)$$

Where;

A = final crash prediction in crashes per year at the railroad crossing

k = normalizing constant (recalculated every two years for passive devices, active devices, and gates)

B = second crash prediction from Equation 2.15

The US DOT formula also includes calculations that determine the probability of a railroad crossing crash being an injury crash or a fatal crash. Every two years, the US DOT recalculates the formula constants based on the most recent five years of crash data. Crash severity is determined by the following equations:[14]

I. The US DOT Crash Severity Equation for Fatal Crashes

$$P(FA|A) = 1/(1 + (\mathrm{KF})(\mathrm{MS})(\mathrm{TT})(\mathrm{TS})(\mathrm{UR})) \quad (2.18)$$

Where;

P(FA|A) = probability of a fatal crash, given a crash

KF = formula constant (440.9)

MS = factor for maximum timetable train speed = $\mathrm{ms}^{-0.9981}$

TT = factor for through trains per day = $(\mathrm{tt}+1)^{-0.0872}$

TS = factor for switch trains per day = $(\mathrm{ts}+1)^{0.0872}$

UR = factor for urban or rural crossing = $e^{0.3571\mathrm{ur}}$

ur = 1 for urban, 0 for rural

II. The US DOT Crash Severity Equation for Casualty Crashes.

$$P(CA|A) = 1/(1 + (\mathrm{KC})(\mathrm{MS})(\mathrm{TK})(\mathrm{UR})) \quad (2.19)$$

Where;

P(CA|A) = probability of a casualty crash, given a crash

KC = formula constant (4.481)

MS = factor for maximum timetable train speed = $\mathrm{ms}^{-0.343}$

TK = factor for number of tracks = $e^{0.1153\mathrm{tk}}$

U = factor for urban or rural crossing = $e^{0.296\mathrm{ur}}$

ur = 1 for urban, 0 for rural

2.4 Regression technique

The number of cars derailed represents non-negative count data, whose mean value can be estimated using regression techniques. Poisson regression and negative binomial (NB) regression are among the most popular count data regression methods used in accident analysis. The Poisson model is suitable for data whose mean is equal to its variance, whereas the NB model assumes that the Poisson mean follows a gamma distribution. The NB model has been used for analyzing over-dispersed data (the variance is greater than mean)[16].

2.4.1 Poisson regression model

In light of the problems associated with linear regression, many turned to Poisson regression as a means to better predict accident frequency:

$$P(y_i) = \frac{e^{(-\lambda_i)}(\lambda_i^{y_i})}{y_{i!}} \quad (2.20)$$

Where $P(y_i)$ for this investigation is the probability of 0, 1, 2, 3…n accidents occurring at the track section i, y_i expected number (0, 1, 2, 3…n) of accidents occurring at track section i, and λ_i is the Poisson parameter defined as:

$$\lambda_i = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_{3+}\dots\dots+\beta_k x_k)} \qquad (2.21)$$

Or more commonly in log-linear form:

$$log\lambda_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_{3+} \dots\dots. +\beta_k x_k \qquad (2.22)$$

Because the Poisson regression model is heteroscedastic, the model coefficients for are typically estimated via maximum likelihood methods. The likelihood function for the Poisson regression model is given as:[17]

$$L(\beta) = \prod_i \frac{e[-e(\beta X_i)][e(\beta X_i)]^{y_i}}{y_i!} \qquad (2.23)$$

One important property of the Poisson distribution is that it restricts the mean and variance of the distribution to be equal:

$$E[Y_i] = Var[Y_i] \qquad (2.24)$$

If this equality does not hold, the data are said to be either under dispersed or over dispersed [18]. Same situation had been discussed by other researchers in their report [19].

$E[Y_i] > Var[Y_i]$ Under dispersed (2.25)

$E[Y_i] < Var[Y_i]$ Over dispersed (2.26)

2.4.2 Negative binomial regression model

Due to common over-dispersion difficulties, use of negative binomial regression techniques was the next evolutionary step in relating accident frequency to various explanatory variables. According to Austin & Carson [18] study, the negative binomial model is more appropriate for over-dispersed data because the model relaxes the constraint of equal mean and variance($E[Y_i] = Var[Y_i]$). This relaxation of the

Poisson constraint is accomplished through the addition of a Gamma-distributed error term to the Poisson model such that

$$\lambda_i = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_{3+}\dots\dots+\beta_k x_{k+\xi})} \tag{2.27}$$

Or in log-linear form:

$$log\lambda_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_{3+} \dots \dots . +\beta_k x_k + \xi \tag{2.28}$$

Where ξ is the Gamma-distributed error term and all other variables are as earlier defined. The addition of ξ either allows the mean to differ from the variance such that

$$Var[Y_i] = E[Y_i][1 + \alpha E[Y_i]] \tag{2.29}$$

Where α is an additional estimable parameter indicative of over-dispersion. If α is not significantly different from zero, the data are not over-dispersed and the Poisson regression is appropriate. The larger the value of α, the more variability there is in the data over and above that associated with the mean λ_i.[17]. Lord & Mannering [20] in their study has mentioned the limitation of Negative Binomial to handle data with small sample size and low sample means value.

2.4.3 Zero-truncated model

Both the Poisson and NB distributions include zeros, so they cannot be directly use to analyze data excluding zero counts, such as the train derailment severity data. The smallest number of cars derailed in a derailment is 1. The Poisson or NB probability functions and their respective log-likelihood functions need to be modified to account for the exclusion of zeros. Compared to the traditional count data models (Poisson or negative binomial), the zero-truncated models calculate the probability of response variable based on positive count data using Bayes's Theorem. The comparison of a ZTNB and NB model shows that the ZTNB model accounts for the exclusion of zeros, thus may have a greater probability and mean value of the response variable, given all else being equal. Below is an equation for ZTNB.

Probability mass function:

$$P_r(Y_i|Y_i > 0) = \frac{(\frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})})\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}(\frac{\mu_i}{\alpha^{-1}+\mu_i})^{y_i}}{1-(1+\alpha\mu_i)^{-\alpha^{-1}}} \quad (2.30)$$

Mean;

$$E(Y_i|Y_i > 0) = \frac{\mu_i}{P_r(y_i>0)} = \frac{\mu_i}{1-(1+\alpha\mu_i)^{-\alpha^{-1}}} \quad (2.31)$$

Variance;

$$Var(Y_i|Y_i > 0) = \frac{E(Y_i|Y_i > 0)}{P_r(Y_i>0)^{\alpha}} = [1 - P_r(Y_i = 0)^{1+\alpha}E(Y_i|Y_i > 0) \quad (2.32)$$

Likelihood function;

$$L = \prod_{i=1}^{N} P_r(Y_i|Y_i > 0) \quad (2.33)$$

$$= \prod_{i=1}^{N} \frac{(\frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})})\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}(\frac{\mu_i}{\alpha^{-1}+\mu_i})^{y_i}}{1-(1+\alpha\mu_i)^{-\alpha^{-1}}} \quad (2.34)$$

Where $P_r(Y_i|Y_i > 0)$ is the probability mass function of zero- truncated negative binomial distribution, $E(Y_i|Y_i > 0)$ is the expectation of zero-truncated negative binomial distribution, $Var(Y_i|Y_i > 0)$ is the variance of zero-truncated negative binomial distribution, α is the over-dispersion parameter, L is the likelihood function, μ_i is the estimated derailment severity for the ith observation, Y_i is the observed derailment severity for the ith observation.[16]

2.4.4 Parameter use in modeling approach

Some of researches [19][21][22] has mention in their research about the parameter required in data analysis to develop prediction model such as below:

I. railway feature: daily trains, train speed, and track number;
II. Highway feature: highway type, highway width, highway grade, highway lanes, highway separation, annual averaged daily traffic, and daily trucks

III. Crossing feature: crossing type, crossing width, crossing length, crossing angle, crossing surface, obstacle detection device, warning bell and flashing light, emergence call, and trolley wire;

IV. Traffic controls: stop line markings, police patrol, and law enforcement camera;

V. Geographic: crash location can be separated by region

VI. Environmental: Lighting and weather condition

All the parameter chosen will be use as independent variable to the prediction model analysis. Usually, this parameter has relation or can be possible causes to the accident and incident. Thus, prediction model is use to evaluate performance of the parameter chosen either good or bad.

2.5 Summary

This literature reviews have discussed several perspectives to predict an accident and incident in railway. In this industry, terminologies of accident and incident is used to refer the whole occurrence happen causing damage of properties, facilities or anything causing human death or injured, or in risk of death or injured. The reviews also have discussed about root cause analysis. This method is importance to use to find out the root cause of problem occur. One of the methods used is Ishikawa diagram. Ishikawa diagram is quality improvement tools to find out the cause effect of problem occurred. There are three elements induced in rail accident, those are human, environment and system. Then, all the possibility of accident can be identify using this tool. Literature review also discuss on previous prediction model used around the world. It starts with Peabody and Dimmick formula in 1941 to predict rail crossing accident frequency. After that, the New Hampshire formula was introduced and become the most optimum formula. This formula is more straightforward with three readily available inputs. Based on this formula, railroad crossings with higher exposure factors or passive warning devices will rank as a higher priority for safety improvements than will railroad crossings with lower exposure factors or more active levels of warning devices. American Association of State Highway Officials (AASHO now AASHTO) and the Association of American Railroads (AAR) in response to the disproportionately high number of accidents occurring at highway-rail crossings has introduce NCHRP Hazard index. This model closely resembles the basic formula of the New Hampshire Index. The others popular prediction model used are USDOT formula in 1980. However, this formula is quite complex compare to the others. Nowadays, most of researcher use regression formula as a method to predict rail accident referring to their characteristic of count data. The Poisson model is suitable for data whose mean is equal to its variance (under dispersion data), while Negative binomial for mean is less than variance (over dispersion data). Furthermore, when the data is excess zero, inflated model is more proper to use. However, when the data are non-negative value, truncated model is more suitable to use. Other than that, this reviews also focusing on what are the parameters used and need to be considered to predict accident or incident model to completing this thesis from previous prediction model.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter, it will briefly discuss on methodology process of prediction assessment. There are several step required before start doing the prediction. Thus, flow process in Figure 3.1 will explain why this assessment is required and important to the railway industry. From that Figure 3.1, it is necessary to conduct some investigation when train incident occurred. However, it is depending on impact from that occurrence. Most of that, the investigation is appeared when it relates to the safety impact or it cause huge damage to the rail system and infrastructure. Thus, counter measure is required to fixing the broken area. Sometimes, upgrading system is needed to improve current system and overcome the problem from repeated. However, the common problem they are facing is a cost limitation. Thus, to ensure the counter measure is effective and applicable with the cost, each contributory factor should have their own priority level. Then comprehensive plan can be set to eliminate those factor one by one. Thus, to realizing it, prediction model was proposed. This prediction model assessment focuses to find the most contributory factor of accident occurrence and prioritizing it. Prediction model also can be use as monitoring system of accident and incident cases. In future, it could help to improve safety feature and reducing number of accident cases in railway.

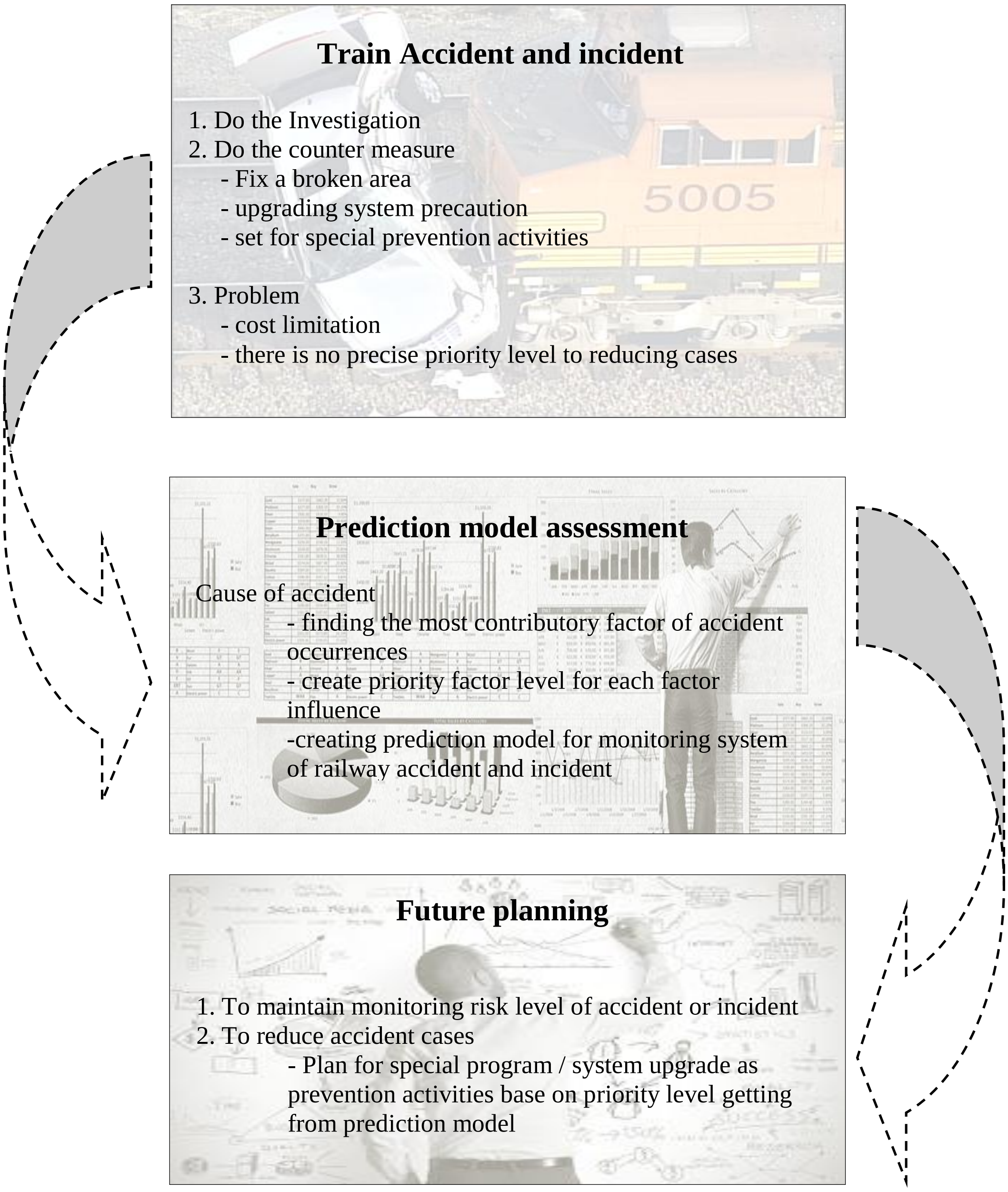

Figure 3.1: Flow of research

In below Figure 3.2, it shows the flowchart of extracting data to analyze the accident and incident data cases getting from investigation report. Then develop prediction equation model using regression formula. This process flow was created to matching the data taken with software tools use to simulate regression formula using R-studio. For further explanation to the process flow, follow below sub topic in Chapter 3.3, 3.4 and 3.5.

REFERENCES

1. Renner, M., Gardner, G., & Alliance, A., Global Competitiveness In The Rail And Transit Industry, *Development*, 2010.

2. Ogden, B. D., Railroad-Highway Grade Crossing Handbook - Revised Second Edition 2007, 2007.

3. Federal Railroad Administration Office of Safety Analysis, Definitions, Retrieved October 2015 , from http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Definitions.aspx

4. NZ Transport Agency Waka Kotahi, Rail Incident / Accident Report Form, Retrieved December 2015, from Http://nzta.govt.nz/assets/resources/rail-incident-accident-form/docs/rail-safety-form.pdf

5. Zarembski, A. M., & Palese, J. W., Managing Risk on the Railway Infrastructure, *Proceedings of the 7th World Congress on Railway Research, Montreal*, 2006.

6. Wang, L. C. , Hazards Identification Model For Rail Rapid Transit Accidents, *J. Mar. Sci. Technology*, 2004, vol. 12(2):78–85.

7. Elzohairy, Y. M., and Benekohal, R. F., Evaluation of Expected Accident Frequency Formulas for Rail-Highway Crossings, 2000.

8. Liu, X., Saat, M. R., & Barkan, C. P. L., Analysis of Major Derailment Causes on Heavy Haul Railways in the United States, *International Heavy Haul Association Conference, New Delhi, India*. 2013.

9. Geoff Vorley, Mini Guide to Root Cause Analysis, *Quality management and quality* , 2008.

10. Hughes R. G., Section VI: Tools for Quality Improvement and Patient Safety, *Patient Saf. Qual. An Evid. -Based Handb. nurses*, 2008, pp. 1–39,.

11. Berg W. D., Evaluation Of Safety At Railroad Highway Grade Crossings In Urban Areas, Purdue University Lafayette Indianna, 1966.

12. Schultz T. G, Evaluation Of Safety At Railroad - Highway Grade Crossing, 1965.

13. Elzohairy, Y. M., & Benekohal, R. F., Evaluation Of Expected Accident Frequency Formulas For Rail-Highway Crossing (No. ITRC FR 98-2,), 2000.

14. Fischhaber P. M., Development Of Light Rail Crossing Specific Crash Prediction Models, University of Colorado,Doctor of Philosophy, 2014.

15. Austin R. D., An Accident Prediction Model For Highway-Rail Interfaces, Montana State University, Master Thesis, 2000.

16. Liu, X., Saat, M. R., Qin, X., & Barkan, C. P., Analysis Of U . S . Freight-Train Derailment Severity Using Zero-Truncated Negative Binomial Regression And Quantile Regression, 2013, vol. 59:87–93,

17. Oh, Jutaek, Simon P. Washington, and Doohee Nam., Accident Prediction Model For Railway-Highway Interfaces, *Accident Analysis & Prevention*, 2006, vol. 38(2):346–56.

18. Austin, R. D. and Carson, J. L., An Alternative Accident Prediction Model For Highway-Rail Interfaces, *Accident Analysis & Prevention*, 2000, vol. 34(1):31–42.

19. Chiou, Yu-Chiun, and Chiang Fu., Modeling Crash Frequency And Severity Using Multinomial-Generalized Poisson Model With Error Components, *Accident Analysis & Prevention*, 2013, vol. 50:73–82.

20. Lord, D., & Mannering, F., The Statistical Analysis Of Crash Frequency Data A Review And Assessment Of Methodological Alternatives, *Transportation Research Part A: Policy and Practice 44.5.* 2010, pp. 291-305.

21. Haleem, Kirolos, and Albert Gan., Contributing Factors Of Crash Injury Severity At Public Highway-Railroad Grade Crossings In The U . S , *Journal of safety research* 53. 2015, pp. 23–29.

22. Hu, Shou-Ren, Chin-Shang Li, and Chi-Kang Lee., Investigation Of Key Factors For Accident Severity At Railroad Grade Crossings By Using A Logit Model, *Saf. Sci.* 2010, vol. 48(2):186–194.

23. The Australian Transport Safety Bureau, “Rail Safety Investigations & Reports, Retrieved December 2015, from https://www.atsb.gov.au/publications/safety-investigation-reports/?mode=Rail.

24. Department of Transport and Main Road Queensland, Safety reports, Retrieved December 2015, from http://www.tmr.qld.gov.au/Safety/Rail-safety/Safety-reports.aspx.

25. Department of Economic Development Jobs Transport and Resources, Rail Incidents, Retrieved December 2015, from http://economicdevelopment.vic.gov.au/about-us/chief-investigator-transport-safety/rail-incidents .

26. The Office of Transport Safety Investigations (OTSI), Rail Safety Investigations, Retrieved December 2015, from http://www.otsi.nsw.gov.au/rail/investigations.php.

27. Chung, M., Correlation Coefficient, pp. 73–74, 2007.

28. Pudprommarat, C., Khamkong, M., & Bookkamana, P., Zero – Inflated Poisson Regression In Road Accidents On Major Road In The North Of Thailand, *IRCMSA Proc*, 2005, pp. 323–330.