ACCIDENT ANALYSIS AND PREVENTION FROM THE CHAIN PERSPECTIVE AND REVISION REPORT

Jinn-Tsai WONG  
Professor  
Institute of Traffic and Transportation  
National Chiao Tung University  
4F, 114 Chung Hsiao W. Rd., Sec. 1, Taipei 100, Taiwan  
Fax: +886 2 2349 4953  
E-mail: jtwong@mail.nctu.edu.tw

Yi-Shih CHUNG  
Ph.D. Candidate  
Institute of Traffic and Transportation  
National Chiao Tung University  
3F, 114 Chung Hsiao W. Rd., Sec. 1, Taipei 100, Taiwan  
Fax: +886 2 2349 4953  
E-mail: yschung.tt93g@nctu.edu.tw

Abstract: One of the most effective approaches interpreting accident causality is from the chain perspective. However, this approach has been applied only to measuring safety improvement for specific countermeasures. This paper emphasizes that the chain concept, which reflects the generating process of an accident, should be borne in mind when analyzing accidents and preventions. A safety framework of the driving behavior is constructed from the chain perspective in support of this argument. Also, a two-stage approach is proposed to adopt the idea in analyzing accident data. The approach includes applying classification techniques at the first stage and causal inference models at the second stage. The relevant methodologies are introduced and possible issues are discussed. It is believed that the proposed framework and approach would help further studies to analyze and interpret accident data in a more thorough perspective.

Key Words: accident analysis, accident prevention, causal chain

1. INTRODUCTION

Transportation professionals and engineers have devoted themselves to exploring the causality of accidents. Understanding the causality of accidents would help us understand not only the occurring process of accidents, but also the possible ways to avoid accidents. However, factual knowledge of traffic accident causality is not easy to come by. Due to the extremely complicated relationships among factors and the restrictions of data and methods (Elvik, 2003, 2006; Hauer, 1997), it has been difficult for researchers to properly build links among affecting factors and accident outcomes. The causes of an accident have usually been described with the closest-to-accident factors. Researchers, however, have tended to analyze accidents with a more thoroughly perspective – looking into not only the accident itself but also the activities and factors prior to and subsequent to the accident (Wong and Chung, 2007). Some accidents were found to be preventable not by correcting driving behaviors but by adjusting behaviors prior to driving (Eby et al., 2000; Simoes, 2003). In other words, an accident may be prevented if one or more undesirable elements during in process were removed (Baker and Ross, 1961; Fleury and Brenac, 2001; Reason, 1997). Therefore, analyzing and preventing accidents from the chain perspective becomes an alternative approach to understanding accident causality.

Analyzing accidents via the chain concept should be taken along with these two elements: the consideration of multiple factors and the ability to make causal inference between factors. The consideration of multiple factors reflects the fact that the generating process of accidents is
complicated. Unless all important factors are accounted for, the confounding-factor effects would bias the estimation results (Elvik, 2002). As for the relationships between factors and accident consequences, there should be directional connection to show their causality. In short, the consideration of multiple factors and the causal relationship between factors are the two required elements in implementing traffic accident analyses and preventions from the chain perspective.

There have been two types of related research that apply to such idea. One of them prescribes the contents of chains. The contents of an accident chain might include the considered factors and the accident outcomes. For example, Elvik (2003) proposed to use a causal chain approach to reduce the possible confounding-factor effects on safety countermeasure evaluation studies. The approach is named with a causal approach since the causality between factors and accident consequences have been designated by professionals and treated as a true causality prior to analyzing the data. The sequence of factors was put into a logical and temporal order; the strength of links between factors was then estimated with data. This chain analysis is particularly useful to evaluate the effectiveness of safety countermeasures related to road improvement since it usually follows physical laws and that the causal relationships between factors are concrete. The second type of research is to explore the possible chains from the data; the plausibility of the possible causality from the outcomes is then judged. Therefore, the causal chains derived from this approach are not limited to the evaluation of safety countermeasures. Instead, all possible causal chains in an accident database can be explored. For example, Chang and Wang (2006) adopted the classification and regression tree technique in analyzing the traffic injury severity in Taiwan. The population and conditions with higher risks of being injured was identified by observing the derived tree. Researches adopting this approach usually interpret the outcomes from the correlation perspective rather than the causality perspective, for the logical and temporal order in the generating process of accidents are not always explicit.

With the becoming comprehensive accident databases, powerful computational capabilities, and mature methodologies, a wider coverage of chains as well as the causal relationships between factors in the chains could both be constructed. That is, the advantages of the aforementioned two types of chain studies are kept; their disadvantages are overcome. Current databases contain rich information. Accidents can be treated as generating from similar causal chains provided that they occur under the conditions with the most identically critical features. The causality of accidents that belongs to similar causal chains could hence be verified with causality inference techniques. In other words, accident chains are revealed by data itself, so the causal chains of interest are not limited. Since the causality of accidents is verified with the causality inference models rather than being directly interpreted from the classification outcomes, the derived causality is more concrete.

The aim of this study is twofold: First is to build a conceptual framework which explains the generating process of accidents from the chain perspective. The necessity and advantages of applying the chain concept on analyzing and preventing accidents will be revealed from the built framework. Second, a two-stage approach is proposed to conduct an accident analysis based on the framework; related issues are discussed as well. The rest of this paper is organized as follows: The conceptual framework is built in Chapter 2; risk factors and their interactions are discussed in Chapter 3; the methodologies and issues to adopt the framework are discussed in Chapter 4 and Chapter 5, respectively. The concluding remarks are made in Chapter 6.
2. BUILDING THE CONCEPTUAL FRAMEWORK OF DRIVING SAFETY

There has been some research proposing frameworks and models to explain driving behavior and its connections between accidents. Few of them were built from the chain perspective but focused on a certain issue (Elvik, 2003; Juarez et al., 2007); others interpreted driving behaviors yet usually put most of their attentions merely on the driving stage (Fuller, 2005; Sümer, 2003; Wilde, 2001). In this study, some of the cases will be extended and integrated as a more general conceptual framework of driving safety.

To understand the causes of an accident, analyzing only the behaviors at the driving stage is not enough. Juarez et al. (2007), for example, proposed a multilevel model to prevent death among minority young drivers from motor vehicle crashes. They suggested that effective prevention should cover the whole driving processes instead of focusing merely on the driving stage. The whole driving process includes the prior-to-driving environment factors, the driving behaviors, and the crash outcomes. In particular, the prior-to-driving environment factors are those which may affect the young driver’s choice on seat belt use or vehicle choice. Fleury and Brenac (2001) also suggested analyzing accidents through looking into the whole driving process. They proposed to analyze accidents at five stages: the situation prior to driving, the driving situation, the discontinuity situation, the emergency situation, and the collision situation. The conditions of one stage are affected by its previous stage and affect its subsequent stage. Both researches indicate that driving behaviors as well as the occurrence of crashes should not be fully determined by local factors, i.e. only factors at the driving stage. Consequently, the construction of the chain framework should be built first from the factors prior to driving until the factors representing the end of the event.

Numerous factors are involved in the chain. Some research proposes to explicitly partition them into several stages such as Fleury and Brenac (2001); other research, however, such as Juarez et al. (2007) and Sümer (2003) who presented a contextual mediated model which divided factors into distal and proximal context, did not. Fleury and Brenac (2001) proposed to divide factors into a distinctive five stages since their approach was proposed to conduct an in-depth study; therefore, detailed and required information for each stage would be collected. On the other hand, Sümer’s approach (2003) was to analyze the relationship among personality, driving behaviors, and accidents. Since the focus was put on linking the connections between psychological factors and resulting driving behaviors, only two levels of connections were represented (i.e. the connection between psychological factors and driving behaviors, and the connection between behaviors and accident outcomes) although psychological factors could affect the activities prior to driving and then affect the driving behaviors. In brief, the partitions of factors along the chain should depend on the available data and the purposes of the analyses. Nonetheless, the clearer sequential connections are the factors, the more solid the results.

It is assumed that our proposed framework is to be adopted in the research with an accident database. An accident database usually consists of three types of data: person, vehicle, and accident characteristics. Although the sequences for all factors can not be fully determined, a rough partition can be achieved. For example, mode choice must be made prior to driving. Therefore, the numerous factors provided by an accident database can be divided at least four stages: prior to driving, driving, incidents or accidents, and rescue (Wong and Chung, 2007). Of which, the rescue stage was not included in the study by Wong and Chung (2007) since their analysis focused on the occurrence of accident types; accident severity was not considered. However, the factors at the rescue stage should be added to complete a

comprehensive framework in order to fully understand the causes of accident severity. Therefore, the occurrences of accidents in our framework will be described with respect to four sequential stages, namely, the prior to driving, the driving, the incident/accident, and the rescue stages.

At the prior-to-driving stage, the decision of the trip characteristics is then the critical factor, for factors affecting safety-related trip characteristics – like when to drive, which route to take, or whether to take passengers or not – should be considered. Elder drivers, for example, are found to develop more driving strategies than youngsters (Eby et al., 2000; Simoes, 2003). The strategies include not driving after dark, reduce going on freeways, driving only in familiar areas, planning routes where protected left turns can be made and driving with a copilot; all of which fit to compensate their physical impairment (Eby et al., 2000; Simoes, 2003). Therefore, the age factor should be represented at this stage. With similar deduction, numerous factors can be found at the prior to driving stage. To organize these factors, the multilevel model proposed by Juarez et al. (2007) is adopted and modified. The trip characteristics are mainly determined by four sets of factors: driver characteristics, vehicle characteristics, local laws and enforcement, and passenger characteristics. Of which, driver characteristics are further affected by social context, national/regional culture, family, and peers; driver and vehicle characteristics are both further affected by public policies such as driver education and the required safety equipments. The necessity of these factors at the driving stage has been declared by Juarez et al. (2007).

The relationships between factors at the driving stage and those at the incident/accident stage have been intensively studied. Some research focused on analyzing individual driving behaviors with respect to behavioral or social sciences such as Wilde (2001), Sümer (2003) and Fuller (2005); other research put the focus on measuring the effects of particular factors on accidents such as traffic flow, surface conditions, enforcement, etc. Although the involved factors are numerous, they can be roughly divided into three types: driver characteristics, vehicle characteristics, and environment factors (Kim et al., 1995). To simplify our framework, all the factors at the driving and the incident/accident stages are represented in these three sets. Of which, the environment factors are further divided into local driving conditions, such as traffic, weather, light and enforcement, and the transportation infrastructure, such as the set up of speed limit, stop signs, surface condition, etc. Detailed discussions of the numerous involved factors and their interactions are given in Chapter 3.

The last stage goes for the rescue stage. The factors at this stage are rarely discussed. The focus would be put on the response of emergency service. Detailed discussions will be given in the subsequent chapter.

All the factors and relationships are illustrated as in Figure 1. The proposed framework is constructed in two dimensions: the time dimension, and the factor interaction dimension. This framework represents that the occurrence of accidents is dynamic, and the factors are interacted at each stage. Moreover, four nodes (three dotted circles and one dotted star labeled “Crash”) are drawn to collect the effect of the interactions resulted from the aforementioned factors. The last dotted star represents the accident outcomes resulted from the accident chain. In addition, the dotted line connecting the five dotted nodes imply that the effects conducted at one stage would accumulate and affect the subsequent stages either immediately, intermediately, or in a long run. In addition to the age factor, another example regarding the vehicle’s characteristics at the prior-to-driving stage is the choice of cars. It is clear that the choice of cars would affect the driving behavior at the driving stage in terms of, for instance,
whether the driver is familiar with the car, and also affect the accident severity at the accident and rescue stage in terms of, for instance, the compatibility of collided vehicles. Obviously, with the proposed framework, the accident generating processes can be more correctly identified and interpreted. Thus, research results based on the framework should be more convincing.

![Figure 1 Conceptual chain framework of traffic accidents](image)

3. RISK FACTOR

Numerous factors have been studied for their relationship with accidents which are not possible to give a complete discussion in a paper with limited length. Therefore, no attempt is made to provide a complete coverage of all possible risk factors. Instead, the aim is to introduce the representative factors and organize them at stages in the proposed framework.

3.1 Prior to Driving

Drivers are going to decide their driving plan at this stage including which route to go, which vehicle to use, what time to start the trip, and expected time to end the trip. These decisions are usually affected by the driver characteristics, vehicle characteristics, passenger characteristics, local laws, and enforcement.

Of the interactions between these factors, the studies related to driver characteristics have grabbed most attentions. For example, different age groups would show different decision characteristics. Older drivers would like to develop strategies such as stopping night driving or finding co-pilots to compensate their declining ability to cope with complex traffic situations (Eby et al., 2000; Simoes, 2003). Yet, young drivers were found relating to alcohol use and seat belt nonuse which would like to increase the accident risks (Ferguson, 2003).

Different enforcement schemes would also affect the driver’s decisions on making trips.
Mountain et al. (2005) claimed that speed management schemes can affect route choice and this can have a significant effect on accidents within the scheme.

For policy factors, the licensing procedures have particularly significant impacts on trip decisions. The licensing procedures for young drivers are concerned since their immature driving skills and they tend to seek risks. The effect of restrain the licensing procedures for young drivers, such as delaying privilege licensure, imposing night driving curfews, and extending periods of supervised practice of driving, have been found positive effect in many areas (Ulmer et al., 2000). Therefore, different policy settings would affect the amount of traffic exposure and the way showing up on roads for different types of drivers.

In sum, although the safety-related studies at the prior to driving stage is not as many as we will see those at the driving stage, it is observed that the decisions of trip characteristics can obviously attribute to the driver characteristics

3.2 Driving

The relationships between the corresponding factors and accidents are reviewed in Subsection 3.2.1 and integrated in Subsection 3.2.2.

3.2.1 Relationships between Factors and Accidents
The risk factors related to the driver characteristics, vehicle characteristics, and passenger characteristics are the first three classes of factors introduced, followed by the factors related to local laws, enforcement, and policy.

1. Driver characteristics

Many factors related to driver characteristics have been considered as connecting to crashes. The socio-demographic factors have been the most intensively studied factors. Of which, age and gender are the two factors which have been particularly extensively studied. Younger drivers are argued to have high rates of crash involvement due to inexperience in assessing traffic situations. The over-representation in accidents for young drivers is partly due to the lack of driving experience (Williamson, 2003); another possibility attributes to young drivers’ risk-taking behavior (Murray, 1997). Yagil (1998) surveyed 693 male drivers in the Israeli army with questionnaire; he found that young drivers are more likely to violate the law than older drivers either from instrumental motives (such as perceived danger of punishment from violations) or from normative motives (such as a sense of obligation to obey the law).

As for the gender factor, some literatures found that male drivers have higher accident rates and result in severer accidents than female drivers do. Male drivers tend to be involved in fatal accidents since their risk-taking behaviors and attitudes; such behaviors include speeding and alcohol consumption. On the contrary, accidents related to female drivers are usually nonfatal due to their immature skills (Massie et al., 1994, 1997; Laapotti et al., 1998). While these observations for male drivers have been consistent in decades, those for female drivers are doubted because of a continuously increasing number of license holders and higher exposure on roads for female drivers than before (Kim, 1995; Forward et al., 1998 and McKenna et al., 1998). Laapotti et al. (2004) claimed that generally, male drivers are risk seeking while female drivers are risk aversion. Moreover, the immaturity in driving skills for female drivers directly relates to possibility and types of accidents although the skill differences between male and female drivers may have declined.
In addition to socio-demographic factors, factors such as psychological and situational factors would affect the occurrence and consequence of accidents as well. Psychological characteristics are very crucial for risk-taking preference and relate to traffic accidents. With observational studies, psychological characteristics are found significantly related to drivers’ socio-demographic factors. Mizell (1997) found the majority of aggressive drivers are relatively young, poorly educated males who have criminal records, histories of violence, and alcohol problems. Shinar and Compton (2004) also found that men were more likely than women to commit aggressive actions. Furthermore, drivers’ psychological characteristics are significant to accidents as well. Beirness et al. (1993) found that the crash-group display a low degree of self-confidence than the non-crash group; Gulian et al. (1989) found that poor self-esteem and high hostility formed a particularly lethal combination.

Situational factors include transient factors and personal habits. The former indicates the factors that may increase risk contributing to states of fatigue, distraction, irritability, and self-doubt (Norris et al., 2000) while the latter refers to personal life habits which may affect the occurrence of accidents such as drinking habits. Moskowitz and Fiorentino (2000) found that the impairment resulted from alcohol consumption include divided attention, drowsiness, decreasing vigilance, increasing reaction time, etc. Most of the studies found that male or younger drivers have significant relationship with alcohol-related accidents (Harrison, 1997; Abdel-Aty et al., 2000; Keall et al., 2004). As for the fatigue factor, it contributes accidents by deteriorating drivers’ alertness, by impairing their judgment, and by slowing their reactions (Lyznicki et al., 1998). As reviewed by Stutts et al. (2003), drivers’ sleep habits and work pattern have been found significantly related to accidents. Night or rotating shift workers and commercial vehicle operators have significant relationships with accidents. Unlike the alcohol and fatigue factors, there is still uncertainty for the contribution of drugs and illness to accidents (Drummer et al., 2004; Hansotia et al., 1991).

2. Vehicle characteristics

In the past, most concerns have been put on the relationships between accident severity and vehicle types as well as the protection equipments. Elvik et al. (1997) found that the overall injury accident rate of heavy vehicles is nearly the same as for passenger cars, but accidents involving heavy vehicles more often result in fatalities or serious injuries than accidents involving passenger cars only. Three fundamental differences between heavy vehicles and passenger cars are found by Abdel-Aty (2004) including: mass incompatibility, stiffness incompatibility, and geometric incompatibility. In particular, the geometry incompatibility, i.e. the imbalance in ride height, would cause significant impact while the collision type is on the frontal (sight reduction) or side (intrusion into smaller vehicles).

Advanced safety vehicle (ASV) has recently become a popular way to avoid accidents. Of which, the installation of intelligent driving support systems aims to help drivers recognize the road environment correctly, warn drivers while errors occur, guide driver’s maneuvering, or to proceed with automatic driving. The main feature of the system is to provide safety-related information to drivers to avoid incidences. Yet, only the right information provided at the right place and at the right time can bring positive effect on reducing accident risks (Wong and Chung, 2007). Inappropriate information style or too much information may cause information overloading or drive drivers to distraction (Yamada and Kuchar, 2006). Moreover, when drivers decrease their speed in response to the warning messages, they tend to raise their following speed to compensate the loss of time (Boyle and Mannering, 2004).
3. Passenger characteristics

The presence of passengers may provide positive effects on accident prevention. Vollrath et al. (2002) found that the presence of passengers could provide a general protective effect; however this is not found for young drivers especially for driving during darkness, in slow traffic and at crossroads.

The seating position of passengers affects the passenger death and injury in traffic crashes. Glass et al. (2000) found that motor vehicle occupants are at a lower risk of death or non-fatal injury when riding in the rear seats of passenger vehicles as compared with riding in the front seat. Similar results are found by Smith et al. (2004).

4. Environment characteristics

Abundant factors related to the environment factors may affect the occurrence of accidents and its consequences. These factors include road design and road furniture, road maintenance, traffic control, weather, and flow conditions. Interested readers can refer to the book by Elvik and Vaa (2004) which gives a very thorough discussion of these factors, except the last two, via systematic overview and meta-analysis.

The weather factors, in addition to their relationships between road factors, may affect drivers’ cognition process. For example, one lagged effect of precipitation over days was discovered in Eisenberg’s research (2004); that is, if it rained a lot yesterday, then on average, today there are fewer crashes. This may come from the adaptive behaviors by drivers.

As for the flow factors, a significant relationship between crashes and mean speed and variation of speed has been found (Garber et al., 2000; Golob et al., 2004). The complexity of information perceived by drivers is higher and the predictability of traffic situation for drivers may be worse while the mean speed and variation of speed increases. Note that this relationship may not be linear since drivers may pay more attention on flow situation while it gets more complicated.

5. Regulation and policy characteristics

The stricter the rule enforcement the more drivers would comply with the rules. However, this is not necessary for all drivers. Yagil (1998) found that the young drivers’ instrumental motives, which are a reaction initiated by a desire to avoid punishment or to receive positive rewards, are weaker than older drivers’.

3.2.2 Integration of Factors

The driving behavior has been characterized by constantly solving problems that involve thinking, choosing and deciding between different alternatives (Vaa, 2001). Several models have been proposed; interested readers can refer to Fuller’s study (2005) for a thorough review.

The Risk Homeostatic Model (RHM) proposed by Wilde (2001) is adopted as the basis in this research to connect other risk factors. Assumed all drivers would have a target level of risk which comes from the perceived costs and benefits of action alternatives. By comparing with the driver’s perceived level of risk, the driver would tend to adjust his driving behaviors to achieve the target risk. The benefits and costs of action alternatives are obtained from either
comparatively risky or safe behaviors. After the driving alternative is taken place, that would be lagged feedback to the driver that may increase or decrease his perceived level of risk or cause an accident. This simple and intuitive structure can accommodate these factors discussed above as illustrated in Figure 2 where the dotted box is RHM.

![Diagram illustrating the interactions between drivers and risky factors.]

The decision making skills are mainly based on the driver’s experience, driver education and physical ability. With more driving experience, the driver is expected to make a decision closer to his desired adjustment more precisely and quickly. Moreover, when with good driver education, the driver is expected to have better sense to make a right decision and thus perform better in decision making skills. This skill is also affected by the driver’s physical ability. For example, the reaction time for a drunk driver is longer.

The vehicle handling skills are affected by the driver’s experience, familiarity with the vehicle, driver education and the vehicle characteristics. With more driving experience, the driver is expected to handle the vehicle better. Yet, this would be affected by his familiarity with the vehicle. The driver may not be able to handle the vehicle well if unfamiliar with the car. Moreover, when with good driver education, the driver is expected to perform better in vehicle handling. The vehicle handling skills are also affected by the vehicle functions. For example, driving a truck is more difficult than driving an automobile.

The perceived level of risk is based on the driver’s perceptual skills and perceived information from the passengers, the vehicle and the environment. The perceptual skills are affected by the driver’s experience, driver education, and physical ability. With more driving...
experience, the driver is expected to be more sensitive to perceive the necessary information. For example, the experienced driver is expected to be able to perceive the necessary information from high speed flow than novice drivers. Moreover, when with good driver education, the driver is expected to be more sensitive to catch important information and thus perform better in perceptual skills. The driver’s physical ability would also affect his perceptual skills such as spatial contrast sensitivity, color perception and visual field.

The driver’s perceived information comes from communications, the vehicle and the environment. When the driver and passengers talk to each other, or the driver uses cell phones, the driver has to handle more complicated information. Moreover, the interaction between the driver and the car is another source of information. Some information is directly revealed from the vehicle equipments such as speedometer, thermometer, etc.; other information comes from the driver’s control and the vehicle’s response such as kinetic energy and friction. Furthermore, information also comes from the environment. The critical information generating from road environment includes horizontal and vertical alignment, degree of curvature, gradient, access control, speed limits, road markings and signs, etc (Proctor et al., 2001). The information tells drivers the road condition and helps drivers adjust their behavior. The weather condition is also important since it would affect the driver’s visual ability and vehicle movement. Therefore, natural light and rain condition is critical for drivers, and wind and snow information for some special areas.

Two types of flow information are critical to drivers: one is flow factors and the other is flow compatibility. As discussed in previous sections, mean speed and variation of speed are two major indexes to accidents. The driver needs to deal with much more information and response more quickly while the flow speed is high and fluctuate considerably. On the other hand, the flow with high mixed types of road users gives more information to drivers than the flow with low mixed types of road users.

The regulatory information reminds and warns the driver to obey the rules; different enforcement schemes provide different information to drivers. For example, the response for drivers may be to slow down the vehicle when seeing an automated photographic speed detector; however, they may also slow down their car to see what happened when seeing the police.

The target level of risk is affected by the driver’s factors. The critical factors include: socio-demographic factors; psychological factors; transitional situational factors; and personal habit.

3.3 Incident/Accident

This stage describes a discontinuous situation within the road safety system such as the driver falls asleep or a sudden stop of the previous car. When a collision happens, a good response of the driver may be able to mitigate the severity. For example, a driver loses the control of the vehicle since the surface is iced; an experienced driver would brake the car gradually rather than immediately. Moreover, he can also take a suitable position to protect himself while a collision happens such as hold his head in the arms. The driver’s action is affected by his experience, driver education and physical ability. With more experience, good driver education and physical ability (e.g. shorter reaction time), the driver is expected to mitigate the severity of the collision and protect himself well.

The protection equipments of a vehicle and the compatibility of collided vehicles would affect
the severity of a collision. The equipments, such as seat belts, airbags and anti-lock brake systems, can protect the driver and occupants to some degree from a collision. On the other hand, the compatibility of the collided vehicles would affect the severity of a collision due to the mass incompatibility, stiffness incompatibility and geometric incompatibility as discussed on the previous section. Those incompatibilities depend on the vehicle types and bumping positions. The severity of a collision can be alleviated when the road design concerns about safety such as installations of safety fencing.

3.4 Rescue

An efficient emergency response provides better service to save the injuries. The efficiency of an emergency response depends on the distance between collision position and service providers, and the flow conditions.

4. METHODOLOGY

The framework has been built and the important risk factors and their interactions have also been introduced. The next step goes to the way to implement the proposed approach; that is, how to proceed with accident analysis and prevention from the very complicated chain perspective.

Recall that there are two steps to implement the proposed approach: First is to classify accidents such that accidents belonging to same classifications are under the condition that most critical features are identical. Second is to verify the causal relationships from some causal inference models. The relevant methodologies are introduced in the following.

4.1 Classification

The classification step is expected to relieve the abundant heterogeneity existing among accidents. Heterogeneity represents the possible presence of unobserved or inattentively accounted driver-, trip-, area-, road-, and other-specific factors (Karlaftis and Tarko, 1998). Unless heterogeneity is appropriately controlled, the estimation results and causality interpretations can be trusted. The adoption of classification techniques can classify accidents into sets with relatively homogeneous attributes. Instead of a whole dataset, sub-datasets are analyzed and less heterogeneity effects are expected.

With the emergence of computational power, the applications of data mining techniques have become very popular including the traffic accident analysis and prevention field. The avoidance of pre-specified functional forms and the ability to simultaneously handle multiple factors may be the two most attractive features to adopt such methodologies (Chang and Wang, 2006). These advantages are particularly useful in adopting the proposed framework since the more the important risk factors are under control, the more homogeneous the results of classifications.

The primary two types of classification techniques in accident analysis are tree-based and rule-based classification techniques. The tree-based techniques are to sequentially break down a whole dataset into smaller and smaller sub-datasets such that the sub-datasets at the deepest nodes are of the least heterogeneity. The sequence of factor loading depends on the choice of classifiers. Common classifiers include entropy, Gini coefficient, Bayesian, etc. The
differences of applying different classifiers on analyzing accidents are usually decided by their prediction accuracy while the entropy classification was popular in earlier research (Vorko and Jović, 2000). On the other hand, a relatively new technique, named classification and regression tree (CART), becomes another popular choice. This technique can automatically search for the best predictors and the best threshold values for all predictors to classify the target variable, and has been shown a useful tool to be able to effectively identify the risky factors affecting injury severity of traffic accidents (Chang and Wang, 2006).

The rule-based technique is another classification technique in traffic accident analysis and prevention. This type of techniques is to learn rules first from a given dataset; thereafter, the accidents in this dataset are classified based on the derived rules. Some common ways to learn rules from a given dataset include Apriori (Geurts, et al., 2003), neural networks (Tseng, et al., 2005), genetic algorithm (Clarke et al., 1998), etc. Recently, the use of the rough set theory becomes another alternative to classify and analyze accidents (Wong and Chung, 2007). Its non-parametric and non-black-box type process enables the theory to become attractive in exploring the features of accident occurrences.

In short, the first step of traffic accident analysis and prevention from the chain perspective is to classify accidents into relatively homogeneous groups with multiple factors. Consequently, each group represents a specific type of accident conditions described by driver characteristics, vehicle characteristics, trip characteristics, environment characteristics, driver’s behaviors, and accident consequences. However, the classification techniques can not identify the sequential relationships between factors which are required to interpret causality. Although Wong and Chung (2007) distinguished factors into four sets with sequential relationships in terms of accident occurrence, factors for each set are numerous. To obtain more accurate causal relationships, another methodology is required.

4.2 Causal Inference

The causality between factors and accident consequences is not easy to verify since most accident analysis and prevention are observational studies rather than experimental studies. Three basic elements are required to claim the causality: 1) Correlation: Cause and effect must vary together; 2) Time sequence: The cause must come before the effect; and 3) Non-spurious: The relationship between cause and effect cannot be explained by any third variable (Pearl, 2000). Studies adopted conventional statistical techniques such as logistic regression are difficult to verify all these elements except the correlations between factors. To overcome these problems, researchers have been proposing many causal inference models. Of which, the model proposed by Pearl (2000) was concerned as a particularly useful tool and has been applied in some traffic accident analysis (Davis and Swenson, 2006).

To construct a causal model, one needs to identify a set of exogenous variables, a set of endogenous variables, and for each endogenous variable a structural equation describing how that variable changes in response to changes in the exogenous and/or other endogenous variables. This possible causality is represented by a directed acyclic graph. Events are defined in terms of values taken on by the model’s variables. Knowledge of these values will almost always be to some degree uncertain. To allow for uncertainty, Pearl (2000) defined a probabilistic causal model as a causal model augmented with a probability distribution over the values taken on by the model’s exogenous variables, so that this probability distribution determines the probabilities to be assigned to the truth or falsity of counterfactual propositions. The probabilities attached to counterfactual statements can be computed by
augmenting the model with nodes reflecting the counterfactual situation, and then applying algorithms for computing Bayesian updates on graphical models.

5. ISSUE

Analyzing accidents from the chain perspective can capture the nature of the traffic accidents; the generation of an accident is coming from a series or a combination of activities. This paper proposes to implement such an idea starting from classifying accidents from an accident database, and then infer the causality for each classification. Although methodologies and databases have been available and continuously improved, the factual knowledge of accident causality is still not easy to come by. Several issues are worthy of consideration.

The first issue is the robustness of the classification results. Each derived classification represents one type of causal chains. Accidents belonging to same causal chains suggest their accident occurrences are similar. In other words, provided that most important risk factors are considered for classifications, accidents coming from same causal chains should be bound together almost surely. Yet, some techniques, such as CART, have relatively unstable classification results; when different adoption strategies, such as stratified random sampling, are applied, the tree structure and the classification accuracy would alter significantly (Chang and Wang, 2006). Therefore, one should be very careful to choose an appropriate classification technique and the adopted strategies.

The subsequent difficulty lies on how to define which factors are important. Analyzing accidents from the chain perspective has the potential to overcome the confounding-factor effects; yet, the researchers should consider most important factors. However, defining the so-called numbers of important factors containing in an accident database is difficult; moreover, factors are not always important for all types of causal chains. A conventional way to select contributing factors is the use of statistical null hypothesis significance testing (NHST for short). NHST has been regarded as a good measure to define the importance of factors. However, a non-significant factor in statistical sense is not equivalent to an unimportant or useless factor in traffic safety (Hauer, 2004). Moreover, the relationships verified in one place may not hold in another place due to the differences of national or regional culture. Consequently, it would be extremely difficult to correctly specify the relationships between factors and accident consequences purely based on literature and professional knowledge. One possible way to relieve this problem is to examine the location of factors in the proposed framework. When a factor locates at earlier stages, such as driver characteristics, it has more potential to be adopted to interpret more types of causal chains. Another way goes to the use of well-behaved data mining techniques such as rough set theory (Wong and Chung, 2007); however, the appropriateness should be further verified.

Different from the issue to select important factors from an existing database, the other issue is to collect vital information which is absent from on-hand databases, especially information of some important indirectly observable or measurable factors. These types of information are not considered in the proposed approach. However, it is possible to collect this required information by experiments.

The next issue is the regression to mean (RTM) phenomenon. When classification is done, each causal chain contains one or numbers of accidents. One causal chain has more accidents than the other should not be immediately claimed that one is a more dangerous condition than
the other. A causal chain with higher number of accidents may reveal by chance. To eliminate the RTM phenomenon, several approaches have been proposed such as statistical quality control or the adoption of empirical Bayes method (Elvik, 2006). However, how to integrate these methods into the proposed framework remains a problem.

The last, but not the least, issue is the determination of the structure for causal inference. In the study by Davis and Swenson (2006), the structure can be pre-determined since their target of interest is rear-end accidents which mainly follow physical laws. Not all causal chains have such explicit sequential relationships. Although some data mining techniques, such as Bayesian networks or EM-algorithms, could help find possible network structure, the plausibility of derived structure requires professional judgments. Therefore, the determination of the causal inference structure is still a problem.

6. CONCLUDING REMARKS

Exploring accident causality is much more difficult than apprehending the correlations among its factors. This paper proposes a framework to understand accident causality from the chain perspective. The becoming comprehensive accident databases, powerful computational capabilities, and mature methodologies provide the opportunities to learn causal chains by doing classifications and applying causal inferences. Moreover, the paper presents a conceptual chain framework of accidents which hopes to become the basis for future implementation of such idea.

The derived causal chains have much potential in practical applications. Since the derived chains contain detailed information about accident occurrences, with such detailed information on-hand, one can estimate the risks faced by drivers by matching their current driver characteristics, trip characteristics, vehicle characteristics, and road and environment characteristics. The individualized, instead of general, safety warning messages, for example, can then be delivered to a certain driver at the matched time and environment.

There are still some data and methodological issues required to be resolved. However, studying accident causality from the chain perspective provides an approach to be closer to accident causality.

REFERENCES


Laapotti, S. and Keskinen, E. (2004) Has the difference in accident patterns between male and
In Proceedings of the 14th ICTCT Workshop, Caserta, Italy.