

Integrated traffic accident database for accident analysis considering driver's accident and violation records

Yasushi Nishida

Research Section, Institute for Traffic Accident Research and Data Analysis (ITARDA),
6-6 Kojimachi, Chiyoda-ku, Tokyo, 102-0083, JAPAN

Abstract

A lot of factors are related to a road traffic accident; particularly human factors such as road use characteristic, driving maneuver characteristic and safety attitude are the major ones. As a random factor is also included, so it is necessary to minimize the contribution of a random factor to identify human factors related to a road traffic accident. There are several standpoints for traffic accident analysis, such as vehicle-based, location-based and driver-based. And it is effective to analyze driver-based traffic accident data for discussion on the relation between human factors and accidents. An integrated traffic accident database system was developed for analysis considering driver's accident and violation records by ITARD, and several studies were carried out for the evaluation. Useful data for discussion on the relation between types of collision and traffic violations, and the effect of accident experience to the following accident were obtained.

INTRODUCTION

Factors related to a road traffic accident are classified into human, road and vehicle factors. Among human factors, road use characteristic and physical driving performance are thought to be related to sex and age, so if these variables are included in accident data, the relation between road use characteristic/physical driving performance and sex/age are able to be discussed without other database. But most of accident data do not include variables related to an individual physical and mental characteristic data, such as safety attitude and driving behavior characteristic, and it is impossible to discuss these topics with accident data only.

Black spot study is a useful technique to identify a dangerous site. But if drivers involved in accidents at a black spot have a common characteristic, for example "an old male driver with a speeding violation record", it is more efficient to focus on such old drivers instead of the spot. Because a benefit of safety measures applied to a road and road facilities is limited to the spot where the measures are implemented, but a benefit of measures applied to old drivers may be expected at other spots of which a road characteristic is similar to the black spot. To discuss these subjects, it is necessary to collect accident data with road user characteristics, accident records and violation records by driver and by spot or road section.

Some drivers are involved in a traffic accident with high frequency, and they are divided into two groups. One is with high-exposure to road traffic and the other is with a high accident rate. And it is hard to analyze a driving characteristic of the latter group without any information about driving characteristics. But if some database with accident and violation records is obtainable, these studies

are possible by statistical analysis of these database.

The effect of a random factor to an accident could be minimized by analyzing plural accidents caused by one driver. And some driving characteristics might be extracted. In addition not only accident characteristics but also relation between type of collision and accident characteristic could be discussed by intervention of accidents caused by drivers with special features. And the same idea could be applied to a traffic violation.

The hypothesis for this study is as follows; an accident or a violation occurs randomly at some rate related with driving characteristics (accident characteristics and violation characteristics), road traffic environment and road use characteristics (Fig.1). Factors of accident characteristics and violation characteristic might be unified as factors of driving characteristics.

OBJECT

It is the main object of this study to develop a new database system for road traffic accident analysis considering driver's traffic accident record and traffic violation record. Other objects are 1) to categorize types of collision and traffic violations for a driver model simulating a mechanism of traffic accident and traffic violation as shown in Figure 1., 2) to develop several methods applied to the developed database system, and 3) to evaluate the usefulness of the developed database system.

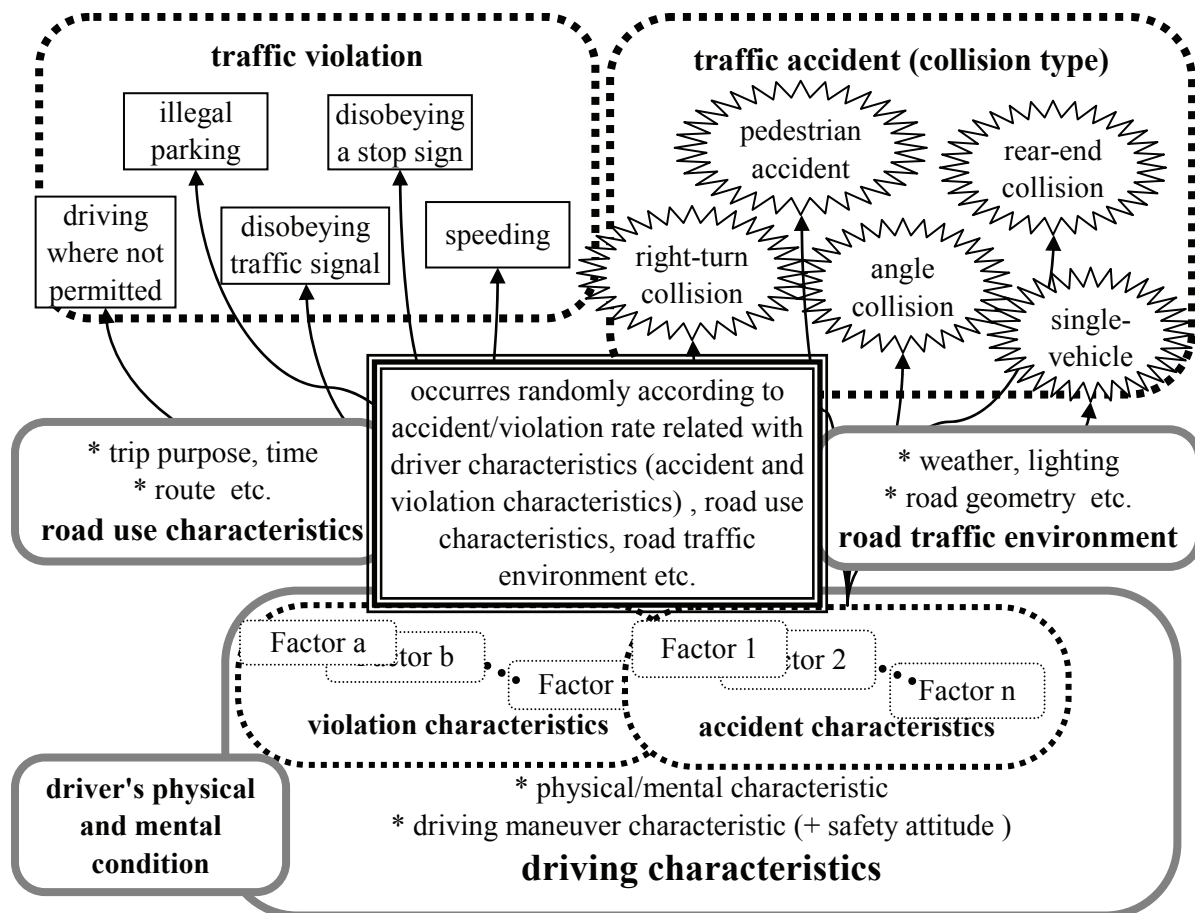


Figure 1. Accident and violation model considering driving characteristics

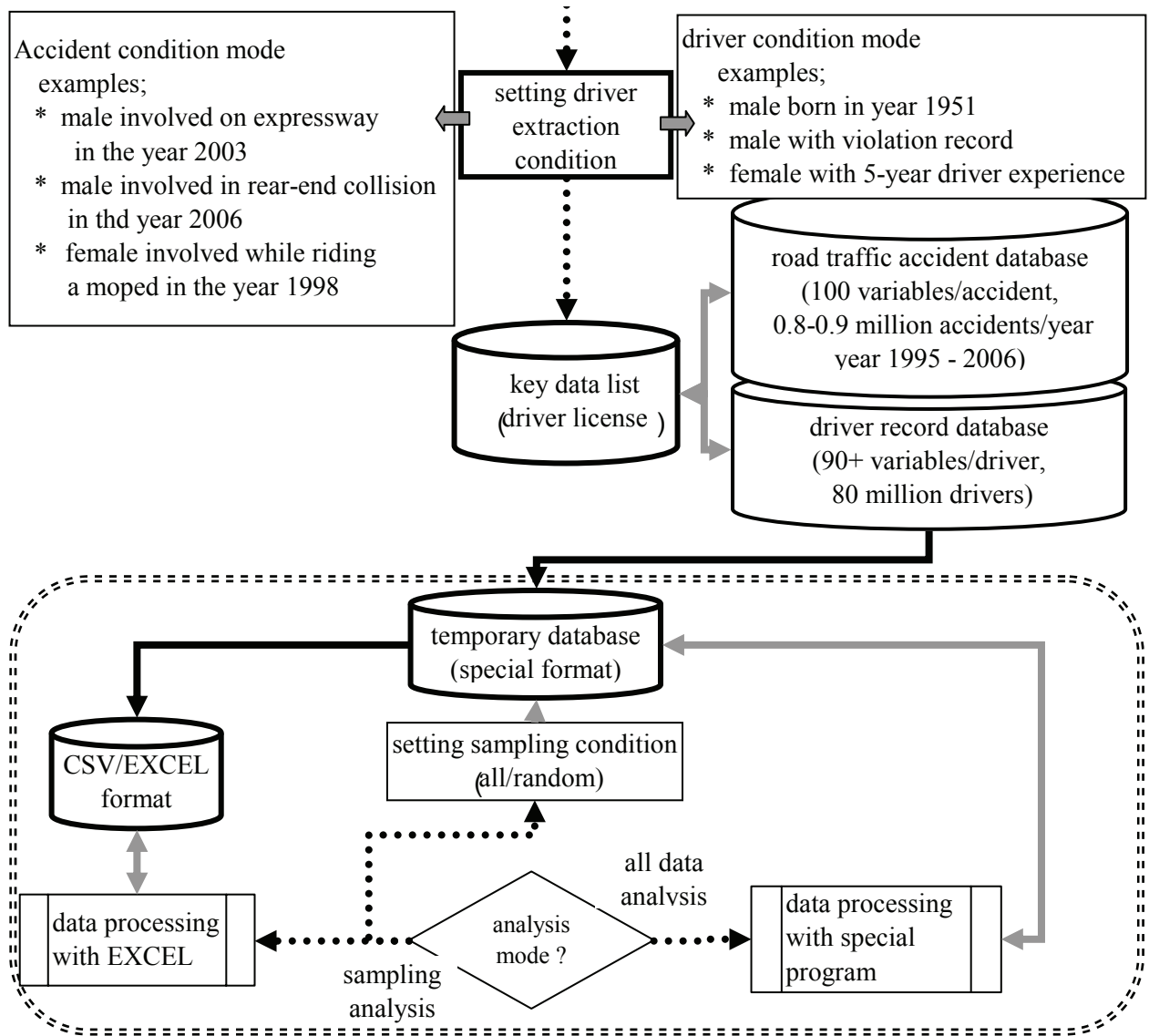
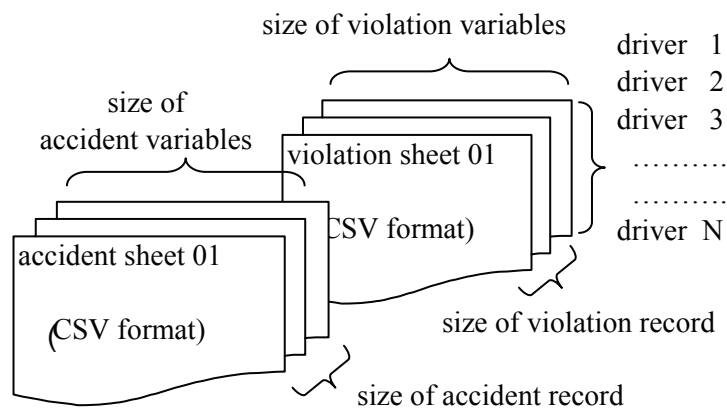


Figure 2. The developed database system



The size of accident/violation record is the maximum value of the driver group concerned.

Figure 3. the structure of the temporary database

DEVELOPED SYSTEM

ITARD (Institute for Traffic Accident Research and Data Analysis) constructed a database by integrating road traffic accident data and road traffic census data in the year 1997 and developed a new database system for analysis considering driver's accident record and violation record in year 2007. The outline of the new database system is shown in Figure 2.

Integration mode

The new database system extracts necessary data from source database by the driver or the accident condition mode. In the both modes, a driver license number is used as key code data for extraction.

Driver condition mode: Key code data is selected with variables of the driver record database, for example, sex, age, city/prefecture, issued date.

Accident condition mode: Key code data is selected with variables of the road traffic accident database, for example, date, type of collision, type of road, age of party, type of vehicle.

Followings are examples of extraction condition;

- 1) drivers involved in road traffic accidents at a designated section of expressway,
- 2) drivers with a speeding violation record,
- 3) drivers involved in a rear-end collision.

Source database

The new database system builds a temporary database extracting required data from the road traffic accident database and the driver record database.

Road traffic accident database: Injured and fatal accidents are recorded, about 100 variables/accident, 0.8-0.9 million accidents/year, and from the year 1995 to 2006.

Driver record database: Driver's general information, accident and violation record are recorded, 90+ variables/driver, 80 million drivers. The size (dimension) of driver record is variable.

Temporary database

The size of integrated database is increased exponentially by integrating two databases, and it is not cost-effective to keep a huge integrated database permanently. So the new database system builds a temporary database according to the object of study. The structure of the temporary database is three dimensional as shown in Figure.3.

Data Analysis

A temporary database is so large that a specific data processing programs is required to analyze all data. But formalized data processing program have not yet be developed, so some of data analyses of this paper are done by EXCEL with data sampled from a temporary database.

RESULTS

Results of example studies using the new database system are shown in this section.

Study 1: Principal Component Analysis for accident and violation characteristics

<Data integrate condition> 943,009 male drivers born in the year 1951

<Sampled data> 56,580 drivers sampled (at 6%) randomly out of 943,009 male drivers

<Analyzed data> Drivers who were involved in a traffic accident or committed a traffic violation, while driving a car or wagon, were analyzed. 7,686 drivers have committed 2 and more traffic violations in the last 5 years, and 347 drivers have 2 and more accident records in the last 5 years.

If plural accidents caused by one drivers are analyzed together, the effect of a random factor is thought to be relatively small and it is possible to discuss accident characteristic or violation characteristic. Data was controlled to reduce the effect of sex, age and type of vehicle in this study. Collision type and traffic violation data of drivers who caused several accidents or violations are respectively analyzed with PCM (Principal component analysis), and Table 1, 2 and 3 show the results. The following explanations for selected vectors are reasonable.

Types of collision (Table 2);

1st Axis: the low level of culpable

2nd Axis: the poor performance of coordinating to other vehicle and road environment

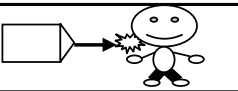
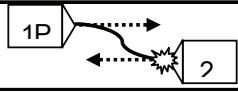
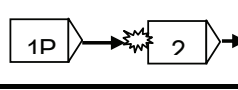
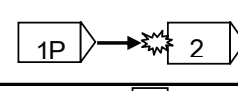
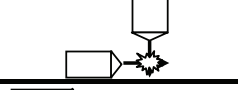

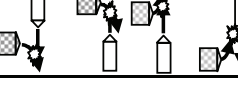


3rd Axis: the lack of courtesy to other road user, especially pedestrian

4th Axis: the high severity of damage

Table 1. Eigenvalues of the correlation matrix

valuables	collision type			variables	traffic violation		
	Eigen value	proportion (%)	cumulative (%)		Eigen value	proportion (%)	cumulative (%)
1	1.376	8.598	8.598	1	1.316	10.123	10.123
2	1.323	8.270	16.869	2	1.226	9.427	19.550
3	1.274	7.962	24.831	3	1.086	8.355	27.906
4	1.182	7.388	32.219	4	1.057	8.133	36.039
5	1.141	7.128	39.347	5	1.037	7.976	44.015
6	1.098	6.861	46.208	6	1.018	7.829	51.844
7	1.072	6.697	52.905	7	1.010	7.771	59.615
8	1.045	6.532	59.437	8	1.000	7.692	67.307
9	0.997	6.232	65.669	9	0.983	7.562	74.869
10	0.958	5.988	71.657	10	0.970	7.458	82.327
11	0.942	5.889	77.546	11	0.955	7.346	89.673
12	0.918	5.740	83.286	12	0.935	7.191	96.864
13	0.869	5.429	88.715	13	0.408	3.136	100.000
14	0.831	5.195	93.910				
15	0.804	5.026	98.936				
16	0.170	1.064	100.000				

Table 2. Axes of PCA for collision types

	collision type	level of culpable	1st Axis	2nd Axis	3rd Axis	4th Axis
	vehicle - pedestrian		-0.341	0.284	0.470	-0.109
	head-on collision	1P	0.036	-0.201	0.207	-0.262
		2P	-0.044	-0.110	0.056	0.439
	rear-end collision with a moving vehicle (rear-end A)	1P	-0.078	0.395	0.246	-0.014
		2P	0.397	0.326	-0.261	-0.027
	rear-end collision with a stopping vehicle (rear-end B)	1P	0.009	-0.679	0.074	-0.299
		2P	0.765	0.249	-0.025	0.064
	collision with a vehicle coming from left or right approach (angle)	1P	-0.265	-0.160	-0.713	0.128
		2P	-0.257	-0.244	0.186	0.287
	collision with an oncoming vehicle while <u>turning right</u>	1P	-0.194	0.084	-0.097	0.156
		2P	-0.056	0.029	0.219	0.592
	collision with a vehicle coming from left or right approach while turning	1P	-0.119	0.033	0.089	-0.532
		2P	0.436	-0.213	0.173	-0.041
	other vehicle-vehicle	1P	-0.208	0.332	-0.291	-0.179
		2P	0.217	-0.182	0.311	0.138
	single vehicle		-0.190	0.350	0.227	-0.071

<level of culpable> 1P: A person having caused the most culpable failure or the least injured among parties concerned when their culpable failure are at the same level. 2P: A person having caused the lower culpable failure.

Table 3. Axes of PCA for traffic violations

traffic violation	1st Axis	2nd Axis	3rd Axis	4th Axis
drunk driving	0.073	0.091	-0.536	0.336
disobeying traffic control signal	-0.083	0.102	0.434	0.302
driving where not permitted	0.100	0.405	0.157	0.108
failing to drive within a designated lane	-0.136	0.129	-0.078	0.159
illegal crossing	0.091	0.317	0.251	-0.206
failing to stop at railway crossing	0.023	-0.177	0.361	-0.150
disobeying a stop sign	0.004	-0.162	0.545	0.425
illegal parking	0.280	0.712	-0.088	-0.171
operating a defective vehicle	0.020	-0.037	-0.018	0.353
speeding	-0.859	-0.160	-0.108	-0.260
failing to use a seat belt	0.662	-0.577	-0.108	-0.147
using a cellular phone while driving	0.076	0.026	0.241	-0.544
unsafely driving	0.074	-0.025	0.014	-0.180

Traffic violation (Table 3);

1st Axis: seatbelt violation or none moving violation

2nd Axis: intentional violation without vicious

3rd Axis: related to recognition error

4th Axis: decision error, optimistic decision

The number of drivers who commit plural accidents or violation is so small that the covariance is not large enough for the PCM. And significant relation between scores of accident vectors and scores of violation vectors was not found (Table 1.).

Study 2: Risk of drivers without accident/violation record

<Data integrate condition> Male drivers born in the year 1930, 1935, 1940, 1945,,, 1975 and 1980

<Sampled data> 40,000 drivers sampled randomly for each group (about 440,000 male drivers)

<Analyzed data> Accident data violation data were analyzed. An active driving-year was estimated to discuss the relation between non-accident/non-violation record and accident involvement.

A road use characteristic may change after retirement and mental and physical performance of driving a car may decrease with age.

There is one scenario for old drivers in Japan. One driver who has neither traffic accident nor traffic violation record may consider himself a safe driver. He may not recognize the declining of his driving performance because he has few near miss or accident. But the reason is not his performance but the low exposure to road traffic (he drives a car weekend only before retirement). But after his retirement an accident risk may increase rapidly because he has enough time to drive and increases exposure and his performance declines with age. Then he may be involved in a traffic accident. It is an accident for him, but it is a reasonable conclusion for society.

Figure 4 shows that the accident involvement in the year 2006 of drivers who have neither accident record nor violation record in the last 3-year (from the year 2003 to 2005). To be a driver without an accident record is not the same, and 50 or 55 years old driver without accident record is more dangerous than 45 years old driver.

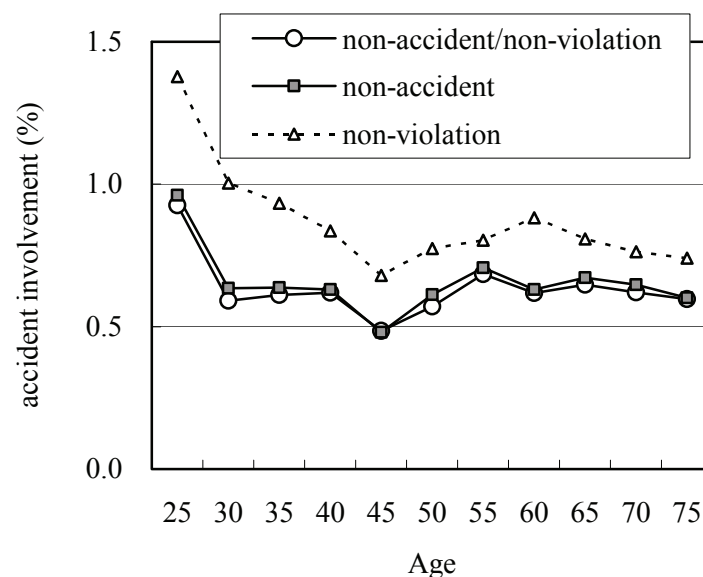


Figure 4. Relation between accident involvement and driver record in the last 3

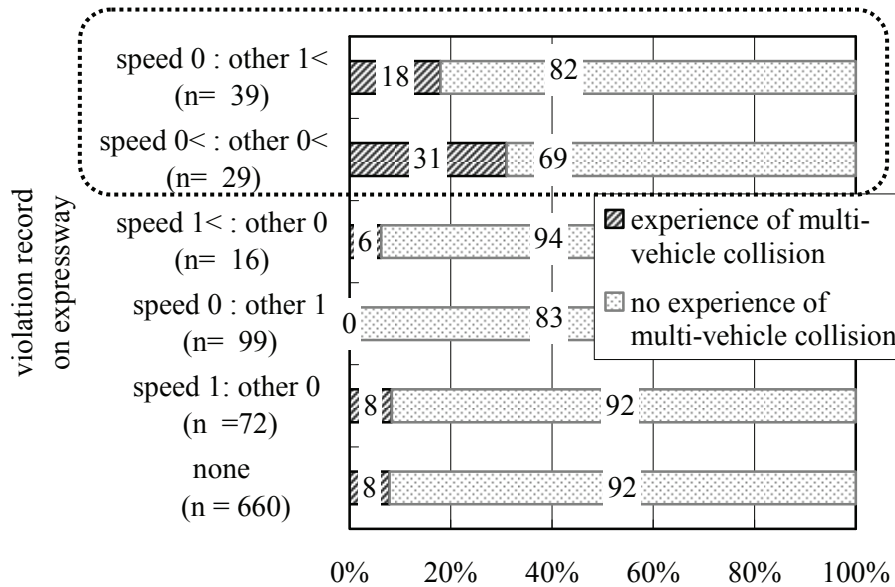


Figure 5. Relation between violation record and multi-collision on expressway

Study 3: Multi-vehicle accident on expressway

<Integrate condition> Drivers involved in traffic accident on Chuoh Expressway in the year 2005

<Analyzed data> All drivers of integrated database (915 drivers) were analyzed. Ninety two drivers out of 915 drivers were involved in a multi-vehicle collision on Chuoh Expressway, and 660 drivers have no violation record on expressways and 117 drivers have speeding violation on expressways.

Multi-vehicle collision on expressway is one of major problem on expressway safety. Major causes of multi-vehicle collision are insufficient headway and speeding, and a driver who commits insufficient headway violation or speeding violation is thought to be involved in a multi-vehicle collision more easily.

Figure 5 shows the relation between violation record and multi-vehicle collision on Chuoh expressway. The result shows that the multi-vehicle collision involvement is high for drivers who have no speeding violation and more than 2 non-speeding violations on expressway. This result is curious, but it is not necessarily unreasonable. A driver may commit a speeding violation when he convinces of his safety, and the correlation between violation and accident is not necessarily strong.

Study 4: The effect of accident experience

<Integrated data> 431,427 male drivers involved in a traffic accident in the year 2006

<Sampled data> 43,142 male drivers sampled randomly (at 10%)

<Analyzed data> Drives involved in a traffic accident while driving a car or wagon were analyzed to eliminate the effect of vehicle type to collision type. Collision type, such as “rear-end B/2P” or “angle/1P” is explained in Table 2.

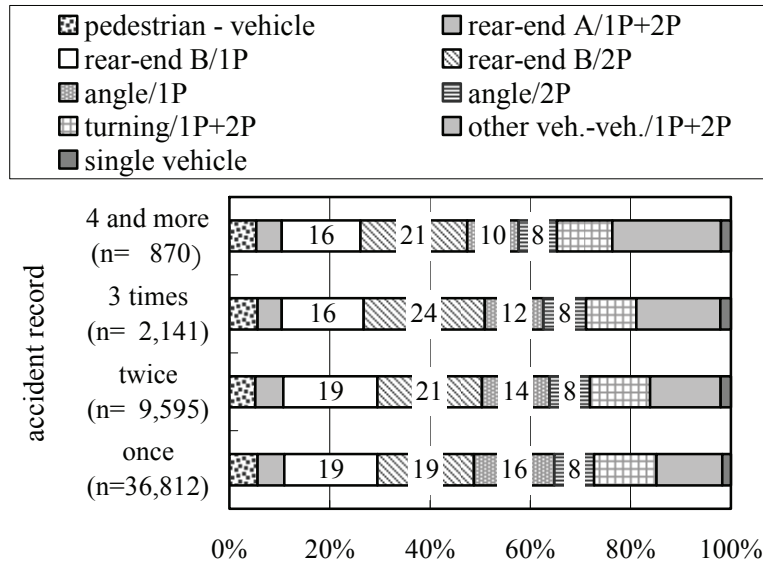


Figure 6. Relation between accident record and collision type

Figure 6 shows the relation between collision type and accident record in the last 5 years. The percentages of rear-end B/1P and angle/1P are decreasing with increase of accident record, but that of rear-endB/2P is increasing (except 4 and more). The result means that an accident characteristic may be changed by an accident experience.

Table 4 shows the effect of collision type to the following accident or sequence of collision type. A percentage of rear-end B/1P is almost 20% and not affected by the type of previous collision, while that of angle/1P is not the same level, especially between 1st and 2nd. The result means that the effect of an accident experience may differ in the collision type.

Table 5 shows that there is a relation between collision type and accident interval. In this analysis, an interval between accidents was controlled, because a driver might pay extra-attention on safety driving just after an accident. An accident occurred in 5 years and longer after the previous accident was selected as a base accident. And only a base accident and the next accident are analyzed for this study. Occurrence rate of rear-end B/2P is thought to be not changed by the driving maneuver of collided drivers, therefore the number of rear-end B/2P could be considered an index of exposure to road traffic. If there is not any effect of driver's safety attitude nor driving maneuver to the next accident, the ratio of the number of collision types concerned to that of rear-end B/2P might be constant. Then using this relation, the number of potential accidents was estimated and shown in Table 5, and accident reductions were estimated by the number of existed accidents dividing by the number of potential accidents. Fig. 8 shows the accident reduction by accident interval.

The result shows;

- 1) a reduction is large just after the accident and getting smaller year by year,
- 2) a reduction of angle/1P is greater than any other type of collision, and
- 3) a reduction is shown for not only primary party (1st party) but also 2nd party.

Table 4. The effect of collision type to the following accident

collision type		rear-end B	angle	rear-end B	angle
sequence of collision type		/1P	/1P	/2P	/2P
Z: number of drivers with 3 straight accident experience while driving a car or wagon		931	931	931	931
A: number of drivers whose 1st accident is the collision type concerned	A/Z(%)	19.4	12.0	21.6	8.1
B: number of drivers whose 1st and 2nd accidents were the collision type concerned	B/A(%)	21.5	8.9	32.3	9.3
C: number of drivers whose 3 accidents were the collision type concerned	C/B(%)	39	10	65	7
D: number of drivers whose 1st and 3rd accidents were the collision type concerned	D/A(%)	23.1	30.0	44.6	28.6
		9	3	29	2
		43	19	64	8
		23.8	17.0	31.8	10.7

Male drivers involved in an accident while driving a car or wagon in the year 2006 were analyzed.

Table 5. Potential accident and existed accident by accident interval

collision type		base accident	accident interval (between a base and the next accident)					
			less than 1 year	1-2 years	2-3 years	3-4 years	4-5 years	5 years and more
Existed accident	rear-end B/1P	4,065	260	250	194	174	149	211
	angle/1P	3,947	204	207	151	133	138	178
	rear-end B/2P	4,503	376	325	237	187	187	225
	angle/2P	1,791	115	114	90	79	73	75
Potential accident	rear-end B/1P	4,065	339	293	214	169	169	203
	angle/1P	3,947	330	285	208	164	164	197
	rear-end B/2P	4,503	376	325	237	187	187	225
	angle/2P	1,791	150	129	94	74	74	89

Accident records of male drivers involved in accidents while driving a car or wagon in the year 2006 were analyzed. A base accidents is an accident that occurred more than 5-year after the previous accident. Collision type: refer to Table 2.

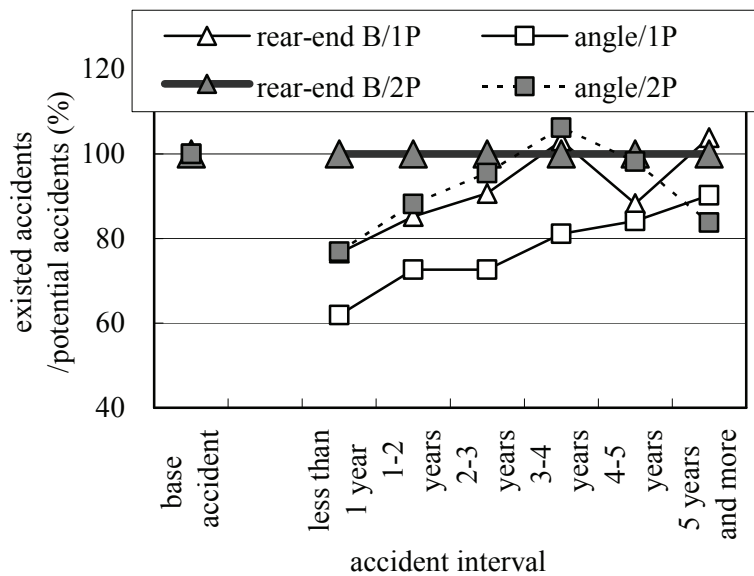


Figure 7. Accident reduction by accident interval

DISCUSSION

Study 1: Principal Component Analysis for accident characteristics

The author tried to derive driving characteristics by studying the relation accident characteristics (Table 2.) and violation characteristics (Table .3). But it was failed because there was not a significant correlation between accident characteristics (scores) and traffic violation characteristics (scores). One of the reasons could be explained by the result of Study 4. The author has supposed that driving characteristics might be so stable that it is possible to derive driving characteristics related to traffic accident and violation. But a driver's attitude and a driving maneuver characteristic may be changed after the driver experiences an accident. Other reason might be the choice of analysis method, and Principal Component Analysis is not suitable method for those data whose correlation coefficient is low. It is required to examine other method and transform those data for the purpose of analysis.

Most of traffic accidents were committed with a human error or unconsciously, but most of traffic violations were committed consciously. So unless the effect of conscious and unconscious behavior is well understood, it is impossible to discuss or construct a driver model for a traffic accident and a traffic violation.

Study 2: Risk of drivers without accident/violation record

A lot of studies showed that 40's is the safest and 20's is the most dangerous age group, and the result of Study 2 is consistent with it. The decreasing of accident involvement at 60's and 70's is thought to be owing to the decreasing of driving exposure of those age groups.

Study 3: Multi-vehicle accident on expressway

Most of traffic accident measures are discussed without considering accidents caused by drivers with accident record. Some of them might try to drive safely and it was shown in Figure 7. There is the difference between the accident rate of drivers with 1 accident record and that of drivers with 2 and more accident records [1]. It means that a counter measures should be changed depending on driver's accident record. Some drivers with a single accident record have ability to control by themselves to prevent an accident, and it is useless to spend a lot of times and effort for them.

Study 4: The effect of accident experience

One of the differences of accident mechanism of rear-end B/IP and angle/IP is the possibility that a driver perceps the other party in advance. It may be preferable for a driver to have such a possibility. But it sometimes has a driver make an error, because "to be visible" is not always "to be perceived".

On the other hand, “not be visible” is always “not to be perceived”. So if a driver intends to avoid an angle collision, he has to reduce his speed and pay attention to others when crossing an intersection. It is not difficult for a driver who has experienced an angle collision once, to avoid another angle collision. A driver assist system to prevent a rear-end collision may be more effect than a driver assist system to prevent an angle collision.

Twenty percents of drivers who were involved in a road traffic accident in the year 2006 have accident record in the last 5 years. It means that if 50% of accidents by those drivers were reduced, 10% of accidents in the year 2006 could have had been reduced.

According to analysis of driver record, only 5 % of drivers are free from a traffic accident or a traffic violation in their lifetime [2]. That is, most of drivers have an experience of accident or violation in their lifetime. And most of these drivers have succeeded to avoid an accident after their unique experience. It is worth analyzing this mechanism.

CONCLUSION

An integrated traffic accident database system for accident analysis considering driver’s accident and violation record was developed. The system was developed to discuss driver characteristics related to both of accident characteristics and violation characteristics, and to obtain useful data for discussion of road safety measures. Unfortunately any driving characteristics related with traffic accident and violation was not obtained, but some useful information was obtained.

The major findings of example studies and points of discussion were as follows;

- 1) an accident risk of drivers without an accident record is changed by aging (Study 2),
- 2) there is no correlation between an involvement of a multi-vehicle collision and speeding violation record on Chuoh expressway (Study 3),
- 3) an accident risk is low just after accident, but a risk is increasing gradually year by year (Study 4),
- 4) these effects are thought to be owing to the change of driver’s safety attitude and driving maneuver characteristic (Study 4),
- 5) it is necessary to examine traffic safety measures from the standpoint of preventing a driver from repeating accident (Discussion), and
- 6) analysis using the developed database system is useful for not only driver education but also for designing driver assist system corresponded to driver’s accident or violation characteristics (Discussion).

REFERENCES

- 1 Repetition of Car Accidents and Traffic Violations, ITARDA INFORMATION No.73, 2008
- 2 Report on Necessary Driving Skills for Safety Driving (2), Japan Safe Driving Center, 2008(in Japanese)