VTI konferens 18A, 2001

Proceedings of the Conference

Road Safety on Three Continents

International Conference in Moscow, Russia,
19–21 September, 2001

Swedish National Road and
Transport Research Institute

VTI konferens 18A · 2001

Proceedings of the Conference

Traffic Safety on Three Continents

International Conference in Moscow, Russia, 19–21 September, 2001

Swedish National Road and Transport Research Institute

Preface

The international conference Traffic Safety on Three Continents in Moscow, 19–21 September 2001, was organised jointly by the Swedish National Road and Transport Research Institute (VTI), the State Scientific and Research Institute of Motor Transport in Moscow (NIIAT), U.S. Transportation Research Board (TRB), the South African Council for Scientific Industrial Research (CSIR), South Africa, and Forum of European Road Safety Research Institutes (FERSI).

The Moscow conference was the 12th in this conference series. Earlier annual conferences have been held in Sweden, Germany, France, the United Kingdom, the Netherlands, Czech Republic, Portugal and South Africa.

Conference sessions covered a number of road traffic safety issues:

- Advanced road safety technology
- Road safety audits
- Policy and programmes
- Traffic engineering
- Vulnerable and old road users
- Alcohol, drugs and enforcement
- Human performance and education
- Behaviour and attention
- Data and models
- Cost and environment
- Speed and speed management

Linköping in November 2001

Kenneth Asp

The influence of sight distance for the speed of vehicles and road safety – Inquiry and comparison in different European countries
Klaus Habermehl, Fachhochschule Darmstadt, Germany

The TRAINER project – development of a new cost-effective Pan-European driver training methodology and how to evaluate it
Torbjörn Falkmer, VTI, Sweden

The effects of diabetes and low blood sugar levels on driving behaviour:comparison of diabetics and non-diabetics.
Marike H. Martens, TNO Human Factors, The Netherlands

The effect of traffic flow improvements on driver attitudes towards pavement markings and other traffic control devices, and pedestrian safety
David Robinson, Fayetteville State University, USA

Session 9. BEHAVIOUR AND ATTENTION

Fatigue of professional truck drivers in simulated driving: A preliminary study
D. Shinar, University of the Negrev, Israel

Dealing with lack of exposure data in road accident analysis
George Yannis, National Technical University of Athene, Greece

Modelling drivers′ behaviour on tapered on-ramps
Antonio D'Andrea, University of Rome "La sapinza",Italy

Driver behaviour models and monitoring of risk: Damasio and the role of emotions
Truls Vaa, TÖI, Norway

Detection and low-cost engineering improvement of inconsistent horizontal curves in rural roads
João Lourenço Cardoso, Laboratório Nacional de Engenharia Civil (LNEC-DVC-NTSR), Portugal

Session 10. DATA AND MODELS

A generic approach for in depth statistical investigation of accident characteristics and causes
Khaled A. Abbas, Egyptian National Institute of Transport, Egypt

A general linear model framework for traffic conflicts at uncontrolled intersections in greater Cairo
Azza M. Saied, Cairo University, Egypt

On the spot accident research in the UK: A new approach to in-depth investigations
Julian Hill, Loughborough University, UK

A GENERIC APPROACH FOR IN DEPTH STATISTICAL INVESTIGATION OF ACCIDENT CHARACTERISTICS AND CAUSES

Dr. Khaled A. Abbas
Egyptian National Institute of Transport
Egyptian Ministry of Transport
+202 2604903

Ahmed T. Al-Hossieny
Transportation Planning Authority
Egyptian Ministry of Transport
+202 2619402

1. INTRODUCTION

The main aim of this research is to develop a generic approach for the utilization of statistical methods to conduct in depth investigation of road accident characteristics and causes. This approach is applied in an effort to analyse the 1998 accident database for the main rural roads in Egypt. This database is composed of traffic accident data collected for 14 road sections representing nine major roads of the Egyptian rural road network.

The proposed approach is composed of two main stages of analysis. Within each stage, several analytical steps are conducted. The first stage is mainly concerned with developing cluster bar charts, where different characteristics and causes of accidents are portrayed in relation to variations in the three main accident contributing factors, namely types of roads, vehicles and drivers. The second stage is concerned with conducting in-depth statistical analysis of the collected accident data. Several parametric and non-parametric statistical tests are carefully selected and applied in accordance with the type of accident data and the objective of the analysis. Within this stage, four levels of statistical investigations were conducted. These are meant to examine a number of issues.

The first level of analysis is concerned with investigating the possible independence of accident characteristics and causes as a result of variations in major governing factors such as types of roads, vehicles or drivers. The main benefit of such procedure is that it can assist in identifying policies and remedial measures that are specific to the accident causes and characteristics of a certain group of roads, vehicles or drivers. In this context two tests are utilized, namely the Chi-square test for categorical data, and the F-test for more than two samples continuous data.

The second level of analysis is meant to compute the appropriate statistical descriptive indicators for accident characteristics and causes. Basically, the mode and median are computed for categorical data, while the mean and standard deviation are computed for continuous data. The third level of analysis is meant to investigate the goodness of fit (conformity) of the collected accident data with generic probabilistic distributions such as Binomial, Uniform, Normal and Poisson distributions.

Finally, the fourth level of investigation is concerned with exploring the association between some selected accident data parameters. In this context, Cramer's V and eta coefficients were computed. The paper presents and discusses the results and conclusions of all of the previous analysis and establishes, in statistical terms, the characteristics and causes of accidents occurring on the Egyptian rural road network.

2. A GENERIC APPROACH FOR IN DEPTH STATISTICAL INVESTIGATION OF ACCIDENT CHARACTERISTICS AND CAUSES

This section is concerned with presenting a generic approach that is developed with the intention of structuring and facilitating an in depth investigation and analysis of accident databases. Several types of in depth investigations of accident data were reported in the literature, see Stoop, 1995, Shankar et. al., 1995, Dionne et. al., 1995 and Grayson and Hakkert, 1987. Most of these research studies were concerned with calibrating plausible explanatory regression models or analysing behavioural issues. This research is concerned with exploring how a logical sequence of statistical analysis could assist in identifying policies and remedial measures that are specific to the accident causes and characteristics of a certain group of roads, vehicles or drivers. In describing the approach, cross-reference to the 1998 accident data base for rural roads in Egypt would be stated. This is meant to demonstrate the practicality of this framework to the analysis of accident data in Egypt. As shown in figure 1, the approach is composed of two main stages.

2.1. First Stage

The first stage is mainly concerned with the descriptive analysis of accident characteristics and causes. Such analysis is conducted by developing cluster bar charts, where different characteristics and causes of accidents are portrayed in relation to variations in the three main accident contributing factors, namely types of roads, vehicles and drivers. The first stage is composed of several steps, see figure 1. The following will provide a brief description of the objective and the analytical procedure followed in each of these steps.

Step 1 is concerned with identifying the most frequently occurring causes of accidents and fatalities. In this step a bar chart is developed showing the frequency of accidents in accordance with the various factors identified as accident causes. The existing system of accident data collection in Egypt allows traffic police officers to mark one or more of 26 listed potential causes of accidents, see ESART, 1999.

Step 2 is concerned with identifying the variability in the frequency of accidents' and/or fatalities' occurrence in accordance with the various types of roads, vehicles and drivers involved in such accidents. In this step, three bar charts are developed to show the frequencies of accidents in accordance with the three well known governing factors of accidents i.e. roads, vehicles and drivers. The latest data collection program in Egypt covers 14 road sections, 13 types of vehicles and four types of drivers (expressed in terms of types of driving licenses).

Step 3 is concerned with reducing the level of the analysis to a manageable level. This step entails the logical aggregation of accident causes. Such aggregation is based on scrutinising the results of step 1 and identifying the governing factors in accidents occurrence. In this context the 26 accident causes were aggregated into 6 main causes, namely road related causes, vehicle related causes, driver related causes, pedestrian related causes, environment related causes and other causes.

Step 4 is also concerned with reducing the level of the analysis to a manageable level. This step entails the logical aggregation of the road, vehicle and driver parameters. Such aggregation is based on scrutinising the results of step 2 and establishing criteria to be used as a basis for such aggregation. Two criterion are proposed, namely the commonality in characteristics and the frequencies of accidents i.e. if there is low frequency in relative terms an aggregation could take place. In this context the 14 road sections were aggregated into 9 main rural roads. Sections 1 through 4 represent the Cairo-Alexandria desert road. Sections 3 and 4 represent the Cairo- Alexandria desert road. Sections 7, 8, 9, 10 and 11 represent Cairo-Fayoum, Cairo-Suez, Cairo- Ismailia, Cairo-Belbeis and Kattamia desert roads respectively. Sections 12 and 13 represent Alexandria-Al Saloum road and finally section 14 represents Sedi Abdel Rahman road. Within the same context the 13 vehicle types were aggregated into 9 types, namely cars, station cars, microbuses, pickups, single truck, truck and trailer, bus, bicycle and motorcycle, and others. As for driver's types, these were kept without any aggregation, i.e. private, first class, second class and third class driving licenses.

Step 5 is concerned with showing the frequency of accidents in accordance with the factors identified as governing accident causes. As mentioned in step 3, six factors were identified as major causes of accidents on rural roads in Egypt.

Step 6 is concerned with developing cluster bar charts. These are used to identify the variability in the frequency of accidents' and/or fatalities' occurrence in accordance with two parameters. The first parameter represents either the variation in accident causes or the variation in one of the accident characteristics. The second parameter represents either the variation in road, vehicle or driver types.

2.2. Second Stage

The second stage is concerned with conducting in-depth statistical analysis of the collected accident data. Each variable in the accident database is identified as either being a nominal (dichotomous or categorical) variable or a cardinal (continuous) variable. Several parametric and non-parametric statistical tests are then carefully selected and applied in accordance with the type of accident data and the objective of the analysis, see Siegel and Castellan, 1988. Within this stage, four levels of statistical investigations are envisaged, see figure 1. These are meant to examine a number of issues, namely the independence of data sets in accordance with variations in key variables, the statistical tendencies, as well as the goodness of fit and association of variables constituting the accident data base.

The first level of analysis is concerned with investigating the possible independence of accident characteristics and causes as a result of variations in major governing factors such as types of roads, vehicles or drivers. In this context three tests can be utilized, namely the Chi-square test for categorical data, the T-test for two samples of continuous data and the F-test for more than two samples of continuous data. Such tests are crucial in determining whether it could be appropriate to conduct further statistical analysis with the whole accident data base or it would be advisable to create branch data bases. Such branch data bases would take into account the variability in accident causes and characteristics resulting from the variability in the groupings of governing factors. In cases where a large number of branch databases is warranted, it could be advisable to conduct further aggregation of governing factors (i.e. reducing the number of groups). Such aggregation would result into more independence results and hence less need for a big number of branch databases. It should be also noted that for some variables, the results may show independence of these variables with one governing factor and no independence with another governing factor. In such cases, priority must be based on the relative importance of governing factors. A logical way is to base such priority in line with the contribution of such factor in accidents occurrence.

The second level of analysis is meant to compute the appropriate statistical descriptive indicators for accident characteristics and causes. Basically, the mode and median are computed for categorical data, while the mean and standard deviation are computed for continuous data. The third level of analysis is concerned with investigating the goodness of fit (conformity) of variables constituting the accident database with generic probabilistic distributions. Several tests are used for such investigation including the Binomial, Poisson for categorical data and Kolmogorov-Smirnov (normal, uniform) for continuous data. Finally, the fourth level of investigation is meant to explore the association between the various collected accident data variables. In this context, a number of association tests including the Phi, Pearson contingency coefficient, and Cramer's V tests can be computed for categorical data, while the Pearson-r test can be used for cardinal data. In case, where association between variables on a nominal and interval scales is warranted the eta coefficient is the most appropriate.

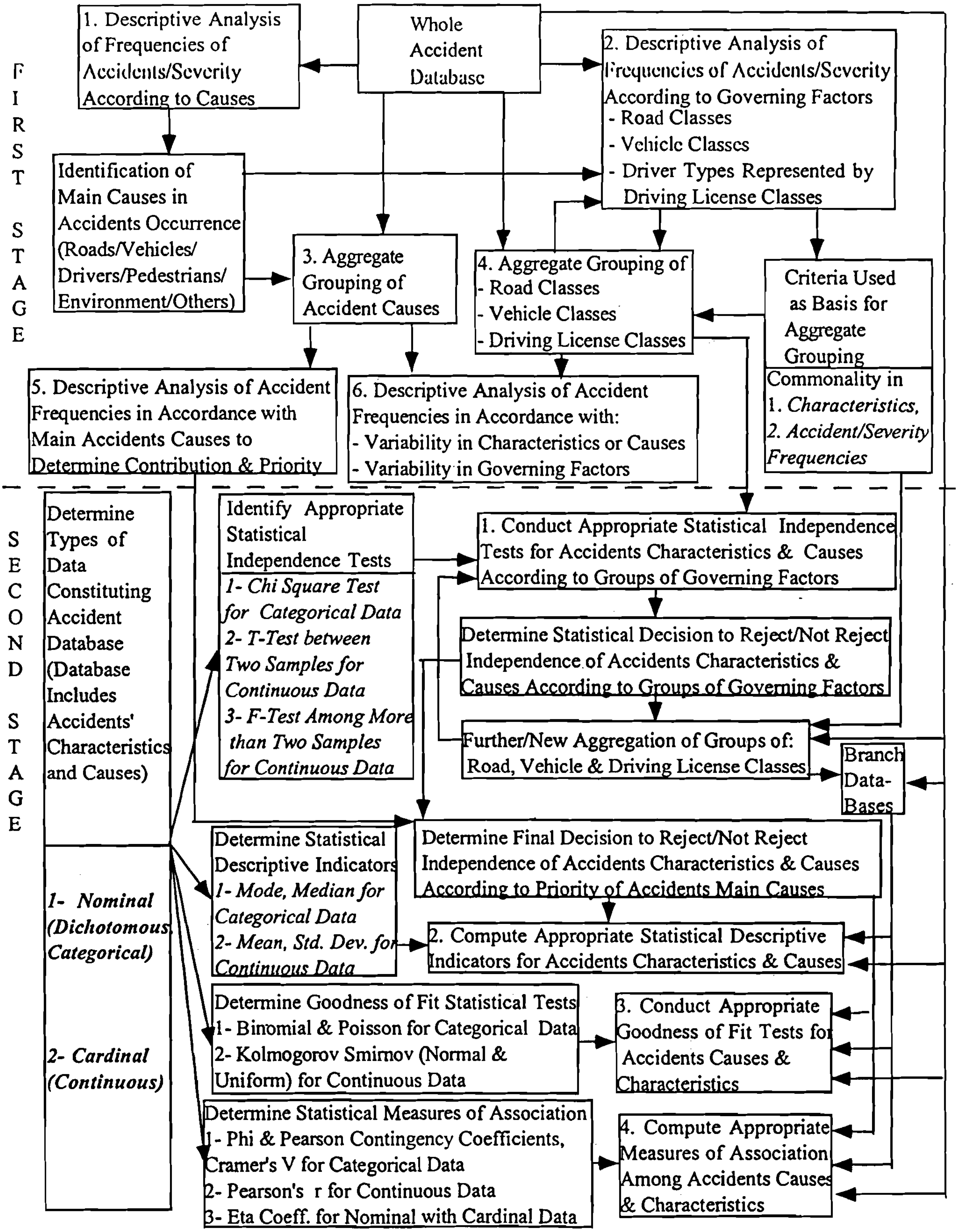

Figure 1: A Generic Approach for In-Depth Statistical Investigation of Accidents' Characteristics and Causes

3. IN DEPTH STATISTICAL INVESTIGATION OF ACCIDENT CHARACTERISTICS AND CAUSES FOR RURAL ROADS IN EGYPT

In the following sections, the second stage of the developed approach will be explored and demonstrated using traffic accident data collected in 1998 for the main rural roads in Egypt. This is meant to achieve an in depth statistical investigation of accident characteristics and causes for rural roads in Egypt. All statistical analysis in this research is conducted using the Statistical Package for Social Sciences SPSS, see Norusis 1993. In this context, it has to be noted that the intention of the paper is not to provide a full comprehensive analysis, but rather a demonstration of the various steps involved. For a more detailed analysis, see Al-Hossieny, 2000.

In stage one of the described approach, conventional descriptive analysis of 1998 traffic accident data, collected on 14 road sections representing nine major roads of the Egyptian rural road network, is conducted. Such analysis is meant to provide a general overview of the patterns, characteristics and causes of accidents occurring on the main rural roads in Egypt. For each accident, data was collected to answer the generic questions of how the accident occurred, when and where did the accident took place, who was involved, and the severity and causes of the accident. Collected accident data was classified as being of the nominal and cardinal type of data. The main outcomes of stage one of the developed approach are as follows.

Three governing factors were identified, namely road related, vehicle related and driver related represented by license types. As mentioned, the 14 considered sections of rural roads were aggregated into 9 roads. Within the same context, the 13 vehicle types were aggregated into 9 types. As for driver's types, these were kept without any aggregation.

The developed accident reporting form contains 27 causes of accidents. These were aggregated into six main causes, namely:

- Driver Related Causes (including 6 causes)
- Pedestrian Related Causes (including 2 causes)
- Vehicle Related Causes (including 12 causes)
- Road Related Causes (including 5 causes)
- Environment Related Causes (including 1 cause)
- Other Causes

Most of the highly contributing causes are driver related. These include loss of control of driving wheel (26.2%), over speed (12.5%), misjudgment of traffic gap (11%), and sudden slowing/stoppage (7.6%). Two other causes related to the vehicle are frequently mentioned, i.e. tire burst and vehicle turnover (16%) or vehicle turn off the road (5%). Together, these six causes contribute around 78.3% of all accident causes on the nine roads. In general driver related causes contribute around 63.9%. This is followed by vehicle related causes contributing in the range of 23.8%. Pedestrian related causes also contribute around 3.8%, while road related causes contribute only 3.6%. Environment and other related causes are also in the range of 3.2%, while other causes are around 1.7%.

4. TESTING INDEPENDENCE OF ACCIDENT CAUSES AND CHARACTERISTICS IN ACCORDANCE WITH GOVERNING FACTORS

The main purpose of this initial analysis is to establish whether accident causes and characteristics would vary with the variation in governing factors. In other words whether as a result of different road characteristics (locations, type, quality), different vehicle characteristics (type, size, load, etc.) or different driver characteristics (education, driving skills, etc.), there will be variations in accident causes and characteristics. Once this is statistically determined, one can proceed with further analysis in one of two directions, see figure 1. The first direction, where no variations are statistically determined, hence the whole accident data set can be used to establish statistical tendencies, goodness of fit and association. The second direction, where variations are statistically determined. In this case, either further aggregation of groups of governing factors can be proposed and re-tested or alternatively proceed with disaggregating the accident database into branch databases. These branch data bases are meant to represent the initial grouping of governing factors. This would create several branch accident databases that can be used to establish statistical tendencies, goodness of fit and association for these groups. The main benefit of such procedure is that it can assist in identifying policies and remedial measures that are specific to the accident causes and characteristics of a certain group of roads, vehicles or drivers.

As previously stated the most appropriate test of independence for categorical data is the Chi square test, while for cardinal data is the F-test in case more than two groups are involved. Results of conducting these tests for all accident characteristics and causes in accordance with the groups of governing factors are shown in table 1. The null hypothesis (H_0) for such tests states that population means of groups of accident causes and characteristics in accordance with governing factors groupings are the same. The table shows that H_0 under 5% significance level is rejected for the main accident causes. This means that it is vitally important to conduct separate investigations and analysis of accident causes for each driver (license) type, vehicle type and specific road. On the other hand, the table shows that H_0, under the same significance level, is not rejected for most accident characteristics. In cases where H_0 is not rejected with one grouping while rejected with another, such as the case with vehicle maneuver, a rule of thumb was thought of. This would establish a priority in the conclusion based on the contribution of governing factors in accident occurrence. Applying such rule, the table shows that for 16 out of the 19 considered accident characteristics, H_0 is not rejected i.e. independence is not rejected. This would lead us to the conclusion that it would be acceptable to compute statistical tendencies, measures of association as well as to test goodness of fit for these 16 characteristics using the entire accident database. The three characteristics, for which H_0 was rejected, are driver age, and driver injury and accident time. For these characteristics, it would be advisable to compute central tendencies, measures of association as well as to test goodness of fit using branch databases disaggregated according to groups of governing factors.

5. ESTABLISHING STATISTICAL CENTERAL TENDENCIES OF ACCIDENT CAUSES AND CHARACTERISTICS

The descriptive statistics in the form of mode and mean (most frequently occurring value) were computed for each of the accident characteristics recorded on nominal and cardinal scale respectively. These are meant to show the central tendencies of these characteristics see table 2. The following conclusions can be deduced:

- Most accidents occur during the month of July. Many summer vacations, leisure journeys and intercity travelling take place during this month. This requires intensification of safety mass media campaigns and enforcement during summer months.
- Most accidents occur on Tuesday. This is a typical working day, where many passenger and freight journeys are undertaken.
- Most accidents occur in dry and clear weather conditions. This suggests that such factors are not detrimental in accident occurrence.
- Most casualties are male casualties. This is expected, as mobility patterns for males are always higher and more intensive than those for females.
- Most pedestrian casualties are fatal. This is also expected as vehicles travel with high speeds on rural roads and pedestrians are physically unprotected and categorised as vulnerable road users. Also, pedestrians usually violate the crossing rules and cross from illegal and unsafe locations. Additionally, the number of protected and well-designed safe crossing locations is limited. The average age of pedestrians involved in accidents is around 34 years. It is vitally important to completely separate pedestrian crossings on rural roads in Egypt. This can be achieved by constructing pedestrian bridges and strictly enforcing the illegal crossings of pedestrians.

Table 1: Testing Independency of Accident Causes and Characteristics in Accordance with Groups of Governing Factors
(Based on 1998 Accident Database for Main Rural Roads in Egypt)

Groups of Governing Factors	Main Causes	Accident Charateristics																		
		Vehicle Maneuver	Accident Date (Months)	Pedestrian Age	Pedestrian Gender	Pedestrian Casualty	Passenger Age	Passenger Gender	Passenger Seat	Passenger Casualty	Driver Age	Driver Gender	Driver Casualty	Damage Level	Collision Type	Accident Location	Road Condition	Weather Condition	Accident Time	Accident Day
	Chi^2	Chi^2	Chi^2	F-test	Chi^2	Chi^2	F-test	Chi^2	Chi^2	Chi^2	F-test	Chi^2	Chi^2	Chi^2	Chi^2	Chi^2	Chi^2	Chi^2	Chi^2	Chi^2
4 License Types	Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	Rej.	N.Rej.	Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	Rej.	N.Rej.
9 Vehicle Types	Rej.	Rej.	Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	Rej.	Rej.	Rej.	N.Rej.	Rej.	Rej.	Rej.	Rej.	N.Rej.	N.Rej.	Rej.	N.Rej.
9 Rural Roads*	Rej.	Rej.	Rej.	N.Rej.	N.Rej.	Rej.	N.Rej.	N.Rej.	Rej.	Rej.	Rej.	N.Rej.	Rej.	Rej.	Rej.	Rej.	Rej.	Rej.	N.Rej.	N.Rej.
Conclusion *	Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	Rej.	N.Rej.	Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	N.Rej.	Rej.	N.Rej.

H_0 : For Chi-Square & ANOVA F-tests H_0 states that population means of groups of accident causes and characteristics in accordance with groups of governing factors are the same.

Rej. = H_0 Rejected i.e. Independence Rejected when Computed Significance Level < 0.05

N.Rej. = H_0 Not Rejected i.e. Independence Not Rejected when Computed Significance Level > 0.05

* Priority Based on Contribution of Governing Factors in Accident Occurrence

Table 2: Descriptive Statistics Showing Central Tendencies of Accident Characteristics

Descriptive Statistics	Accident Main Causes	Vehicle Maneuver	Accident Date (Months)	Pedestrian Age	Pedestrian Gender	Pedestrian Casualty	Passenger Age	Passenger Gender	Passenger Seat	Passenger Casualty	Driver Age	Driver Gender	Driver Casualty	Damage Level	Collision Type	Accident Location	Road Condition	Weather Condition	Accident Time	Accident Day
Mode	*	Moving	July	N.A.	Male	Death	N.A.	Male	Back Seat	Slight	N.A.	Male	**	Major	Front	Along the Road	Dry	Clear	**	Tuesday
Mean	N.A.	N.A.	N.A.	34.4	N.A.	N.A.	31.8	N.A.	N.A.	N.A.	**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

* See Table 3　　** See Table 4　　N.A. : Not Applicable

- On the other hand, most driver and passenger casualties range from slight to severe injuries. These road users are relatively protected in their vehicles. The average age of passengers involved in accidents is around 32 years.
- Most of the passengers involved in accidents are seated at the back of carrying vehicles. Many of the accidents occurring are for cars and microbuses. In Egypt, intercity trips using cars are characterized by high occupancy, as families travel together to spend vacations. Also all passengers involved in microbus accidents are considered as back seated except for the driver and the front seated passenger. Current traffic law in Egypt requires front seated drivers and passengers to wear seat belts. It would be highly desirable to include all back seated passengers to put on seat belts. In addition, the process of conducting driving tests and issuing driving licenses should be reviewed with the objective of ensuring safe driving skills and behaviours.
- Most of the accidents occurred to vehicles while moving along the road. This is typical of rural road accidents. Most accidents are also front collisions.

As for accident causes, it was concluded that it is better to establish the patterns of accident causes for each group of governing factors. In this context, table 3 shows the numeric and percentage contribution of each of the six considered accident causes for the 14 considered sections of rural roads. It is obvious from the analysis that for all road sections, driver related factors are the dominant causes for accidents with the exception of sections 10 and 14 where vehicle related causes are more dominant.

Central tendencies for the three accident characteristics, for which H_0 for the independence test was rejected, were computed for each group of driver (license) types, see table 4. The table shows that the average age of drivers possessing first class licenses varies slightly from other drivers involved in accidents. The table also shows that most first class license drivers, involved in accidents, suffer from serious injuries, while for other drivers involved in accidents, they suffer from slight injuries. Finally the table also demonstrates that most accidents involving drivers with first class licenses occur around 7.00 A.M., while those accidents involving other drivers occur around 8.00 A.M. These results are vitally important in pointing out the relatively high risk of first class drivers and that specific remedial measures to reduce the possibility and severity of accidents involving first class drivers should be thought of.

6. TESTING GOODNESS OF FIT FOR ACCIDENT CAUSES AND CHARACTERISTICS

An important step in accident analysis is to establish how well the data fits well-known probabilistic distributions. Generally, there are two types of probabilistic distributions. The first type include the well known continuous probability distributions such as the normal, uniform, T and F distributions. The second type includes the discrete probability distributions such as the Binomial and Poisson distributions. Characteristics of such distributions are well documented in the literature.

For each of the accident characteristics, the proper distributions are selected in accordance with the type of data i.e. goodness of fit to continuous probability distributions would be tested for cardinal data. On the other hand, goodness of fit to discrete probability distributions would be tested for categorical data. In all goodness of fit tests, the H_0 states that the tested data is in conformity with a specific selected distribution and its assumptions. Results of computing these tests show that for most accident characteristics H_0 under 5% significance level is rejected i.e. none of the considered distributions can describe the data. This suggests a high degree of randomness in the data. Few exceptions were identified, where H_0 was not rejected when attempting to describe the age of pedestrian victims using the normal distribution. In addition H_0 was also not rejected when attempting to describe the gender of pedestrian victims using the Poisson distribution.

Table 3 : Numeric and Percentage Share of Accidents Main Causes in Accordance with Considered Road Sections

Roads	Causes	User Related Causes (Driver)	User Related Causes (Pedestrian)	Vehicle Related Causes	Road Related Causes	Environment Related Causes	Other Causes	All Causes
Road 1	Number	76	18	30	5	9	6	144
	Percentage	52.8%	12.5%	20.8%	3.5%	6.3%	4.2%	100.0%
Road 2	Number	64	5	21	2	2	2	96
	Percentage	66.7%	5.2%	21.9%	2.1%	2.1%	2.1%	100.0%
Road 3	Number	103	4	16	5	10	2	140
	Percentage	73.6%	2.9%	11.4%	3.6%	7.1%	1.4%	100.0%
Road 4	Number	96	7	11	6	0	2	122
	Percentage	78.7%	5.7%	9.0%	4.9%	0.0%	1.6%	100.0%
Road 5	Number	123	3	46	3	0	2	177
	Percentage	69.5%	1.7%	26.0%	1.7%	0.0%	1.1%	100.0%
Road 6	Number	60	2	9	0	0	1	72
	Percentage	83.3%	2.8%	12.5%	0.0%	0.0%	1.4%	100.0%
Road 7	Number	72	4	40	5	0	1	122
	Percentage	59.0%	3.3%	32.8%	4.1%	0.0%	0.8%	100.0%
Road 8	Number	95	3	49	8	4	3	162
	Percentage	58.6%	1.9%	30.2%	4.9%	2.5%	1.9%	100.0%
Road 9	Number	225	14	93	15	16	4	367
	Percentage	61.3%	3.8%	25.3%	4.1%	4.4%	1.1%	100.0%
Road 10	Number	10	0	15	1	0	0	26
	Percentage	38.5%	0.0%	57.7%	3.8%	0.0%	0.0%	100.0%
Road 11	Number	28	1	6	2	1	2	40
	Percentage	70.0%	2.5%	15.0%	5.0%	2.5%	5.0%	100.0%
Road 12	Number	50	1	20	6	7	0	84
	Percentage	59.5%	1.2%	23.8%	7.1%	8.3%	0.0%	100.0%
Road 13	Number	23	0	13	1	3	2	42
	Percentage	54.8%	0.0%	31.0%	2.4%	7.1%	4.8%	100.0%
Road 14	Number	10	0	16	0	0	0	26
	Percentage	38.5%	0.0%	61.5%	0.0%	0.0%	0.0%	100.0%
All Roads	Number	1035	62	385	59	52	27	1620
	Percentage	63.9%	3.8%	23.8%	3.6%	3.2%	1.7%	100.0%

Table 4: Descriptive Statistics Showing Central Tendencies of Accident Characteristics in Accordance with Grouping of Drivers' (License) Types

Accident Characteristics	Descriptive Statistics	Drivers' (License) Types			
		First	Second	Third	Private
Driver Age	Mean	39.35	36.14	36.82	36.67
Driver Casualty	Mode	Serious	Slight	Slight	Slight
Accident Time	Mode	07:00	08:00	08:00	08:00

7. COMPUTING MEASURES OF ASSOCIATION AMONG ACCIDENT CAUSES AND CHARACTERISTICS

In this section, indexes that attempt to quantify the relationship between variables in a cross-classification are called measures of association. No single measure adequately summarises all possible types of association. In this research, three types of measures of association are suggested. The first is meant to establish the association between two categorical accident variables. The second is meant to establish the association between two continuous accident variables. The last is meant to establish the association between two accident variables where one is of categorical type while the other is of continuous type.

In this context, table 5 shows a diagonally symmetric matrix, where rows and columns represent the various accident characteristics. Cells of this matrix show whether it is logically possible to have an association between two characteristics or not. For some cells, where it is plausible to assume associations, appropriate measures are computed. For categorical data, Cramer's V is computed, while for association between categorical and continuous data eta coefficient is computed. Values for both coefficients range from 0 to 1, where 0 represents no association, while 1 represents high association.

Cramer's V values for four associations were in the range of 0.2 to 0.34. These demonstrate the existence of association between type of pedestrian casualty and vehicle maneuvers (0.34) as well as between type of pedestrian casualty and collision type (0.27). In addition, associations were shown between damage level and collision type (0.2) as well as between collision type and vehicle maneuver (0.215). Three strong associations were also established by computing eta coefficients. The first is an association demonstrating the possible dependency of type of pedestrian casualty on pedestrian age (0.7). The second shows the possible dependency of type of passenger casualty on passenger age (0.275). The third also shows the possible dependency of type of driver casualty on driver age (0.28).

In table 6, association measures between level of vehicle damage and type of driver casualty are computed in accordance with groups of drivers' (license) types. The Pearson contingency coefficient and Cramer's V for drivers possessing first class licenses are higher than the values for other types of drivers. This suggests a strong association between level of vehicle damage and type of driver casualty for those drivers possessing first class license and involved in accidents on the rural roads in Egypt.

Table 5: Computation of Some Possible Logical Associations Between Accident Causes and Characteristics

		Accident Characteristics																	
	Accident Main Causes	Vehicle Maneuver	Accident Date (Months)	Pedestrian Age	Pedestrian Gender	Pedestrian Casualty	Passenger Age	Passenger Gender	Passenger Seat	Passenger Casualty	Driver Age	Driver Gender	Driver Casualty	Damage Level	Collision Type	Accident Location	Road Condition	Weather Condition	Accident Time	Accident Day
Accident Main Causes	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Vehicle Maneuver		-	X	X	X	0.342	X	X	X	✓	✓	✓	✓	0.143	0.215	✓	✓	✓	X	X
Accident Date (Months)			-	X	✓	X	X	X	X	X	X	X	X	X	X	X	X	0.174	X	X
Pedestrian Age				-	X	0.703	X	X	X	X	X	X	X	✓	✓	X	X	X	X	X
Pedestrian Gender					-	0.088	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Pedestrian Casualty						-	X	X	X	X	✓	✓	X	✓	0.270	✓	✓	✓	✓	✓
Passenger Age							-	✓	✓	0.275	X	X	X	X	X	X	X	X	X	X
Passenger Gender								-	0.124	✓	X	X	X	X	X	X	X	X	X	X
Passenger Seat									-	✓	X	X	X	✓	X	X	X	X	X	X
Passenger Casualty										-	X	X	X	✓	0.095	✓	✓	✓	✓	✓
Driver Age											-	✓	0.275	✓	✓	X	X	X	X	X
Driver Gender												-	0.065	X	✓	X	X	X	✓	X
Driver Casualty													-	✓	✓	✓	✓	✓	✓	✓
Damage Level														-	0.201	✓	✓	✓	✓	✓
Collision Type															-	0.071	✓	✓	✓	✓
Accident Location																-	X	X	X	X
Road Condition																	-	0.654	X	X
Weather Condition																		-	0.285	✓
Accident Time																			-	0.284
Accident Day																				-

X : No Logical Association

✓ :Possible Logical Association

For Association between Nominal data Cramar's V is computed

For Association between Nominal and Continous data Eta Corfficient is computed

Table 6: Association Measures Between Level of Vehicle Damage and Type of Driver Casualty in Accordance with Grouping of Drivers' (License) Types

Association Measures	Drivers' (License) Types			
	First	Second	Third	Private
Pearson Contingency Coefficient	0.412	0.348	0.351	0.354
Cramer's V	0.319	0.263	0.265	0.267

8. CONCLUSIONS

The main aim of this research was to develop a generic approach for the utilization of statistical methods to structure and facilitate the in depth investigation of road accident characteristics and causes. This approach was applied in an effort to analyse the 1998 accident database for the main rural roads in Egypt. This database is composed of traffic accident data collected for 14 road sections representing nine major roads of the Egyptian rural road network. The research explored how a logical sequence of statistical analysis could assist in identifying policies and remedial measures that are specific to the accident causes and characteristics of a certain group of roads, vehicles or drivers.

The proposed approach is composed of two main stages of analysis. Within each stage, several analytical steps are conducted. The first stage is mainly concerned with developing cluster bar charts, where different characteristics and causes of accidents are portrayed in relation to variations in the three main accident contributing factors, namely types of roads, vehicles and drivers. The second stage is concerned with conducting in-depth statistical analysis of the collected accident data. Several parametric and non-parametric statistical tests are carefully selected and applied in accordance with the type of accident data and the objective of the analysis. This paper was mainly concerned with presenting and demonstrating the second stage of the developed approach Within this stage, four levels of statistical investigations were conducted.

The first level of analysis is concerned with investigating the possible independence of accident characteristics and causes as a result of variations in major governing factors such as types of roads, vehicles or drivers. The second level of analysis is meant to compute the appropriate statistical descriptive indicators for accident characteristics and causes. The third level of analysis is meant to investigate the goodness of fit (conformity) of the collected accident data with generic probabilistic distributions. Finally, the fourth level of investigation is concerned with exploring the association between some selected accident data parameters. The following presents a summary of conclusions:

Independencies of main accident causes in accordance with groups of roads, vehicles and drivers were rejected. This means that it is vitally important to conduct separate investigations and analysis of accident causes for each driver (license) type, vehicle type and specific road. Further analysis showed that for all road sections, driver related factors are the dominant causes for accidents with the exception of sections 10 and 14 where vehicle related causes are more dominant. On the other hand, independence is not rejected for most accident characteristics (16 out of 19). The three characteristics, for which independence was rejected, are driver age, and driver injury and accident time. Average age of drivers involved in accidents slightly vary with the variation in types of licenses. First class licensed drivers, involved in accidents, suffer from serious injuries, while for other drivers involved in accidents, they suffer from slight injuries. Most accidents involving drivers with first class licenses occur around 7.00 A.M., while those accidents involving other drivers occur around 8.00 A.M. These results are vitally important in pointing out the relatively high risk of first class drivers and that specific measures to reduce the possibility and severity of accidents involving first class drivers should be thought of.

Most accidents occurred during the month of July, on Tuesdays, and in dry and clear weather conditions. Most accidents occurred to vehicles while moving along the road. This is typical of rural road accidents. Most accidents are also front collisions.

Most casualties are male casualties, where most pedestrian casualties are fatal with an average age of 34 years. On the other hand, most driver and passenger casualties range from slight to severe injuries. The average age of passengers involved in accidents is around 32 years. These are mostly back seated

In testing conformity of accident data with selected probabilistic distribution, results showed that for most accident characteristics, none of the considered distributions can describe the data. This suggests a high degree of randomness in the data. Few exceptions were identified, where the age of pedestrian victims can be described using the normal distribution, while the gender of pedestrian victims could be described using the Poisson distribution.

Several logical associations among accident data were identified and some were computed such as the association between type of pedestrian casualty with vehicle maneuvers as well as with collision type. In addition, associations were shown between damage level and collision type as well as between collision type and vehicle maneuver. Associations were also established demonstrating the possible dependency of type of casualty on age of casualty. Association measures between level of vehicle damage and type of driver casualty were computed in accordance with groups of drivers' (license) types. The Pearson contingency coefficient and Cramer's V for drivers possessing first class licenses proved to be higher than the values for other types of drivers. This suggests a strong association between level of vehicle damage and type of driver casualty for those drivers possessing first class license and involved in accidents on the rural roads in Egypt.

REFERENCES

Al-Hossieny A. T. (2000) Development of a Generic Approach for Utilisation of Statistical Methods to Conduct In Depth Investigation of Road Accident Characteristics & Causes: A Case Study of Main Rural Roads in Egypt. MSc Thesis. University of Westminster, UK.

Dionne G., Desjardins D., Laberge-Nadeau C., and Maag U. (1995). Medical conditions, risk exposure, and truck drivers' accidents: An analysis with count data regression models. Accident Analysis and Prevention, Volume 27(3), pp. 295-305. Pergamon Press.

Egypt Academy for Scientific Research and Technology (EASRT) (1999) Investigation of traffic accidents on intercity roads in Egypt. Final Report Conducted for the Transport Research Council, by the Egyptian Traffic Police Department and Academics.

Grayson G. B. and Hakkert A. S. (1987) Accident analysis and conflict behaviour. In: Road Users and Traffic Safety. Eds. J. A. Rothengatter and R. A. de Bruin Van Gorcum. Assen

Norusis M. J. (1993) Statistical Package for Social Sciences (SPSS) for Windows. Base System User Guide, Release 6.0. Spss Inc., USA.

Shankar V., Mannering F. and Barfield W. (1995) Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention, Volume 27(3), pp. 371-389. Pergamon Press

Siegel S., and Castellan N. J. (1988) Nonparametric statistics for the behavioural sciences. McGraw-Hill.

Stoop J. A. (1995) Accidents – In - Depth Analysis; towards a method AIDA?. Safety Science, Volume 19, pp. 125-136. Elsevier Press.