UNIVERSITY OF CATANIA

Faculty of Engineering

Department of Civil and Environmental Engineering

PhD Course on

"Engineering of Hydraulic Infrastructure, Environmental Health and Transportation"

CYCLE XXVI

A STOCHASTIC APPROACH TO SAFETY MANAGEMENT OF ROADWAY SEGMENT

CARMELO D'AGOSTINO

Supervisor:

Professor Salvatore Cafiso – University of Catania, Catania (IT)

Co-Supervisor:

Professor Bhagwant Persaud – Ryerson University, Toronto (CA)

Research Doctorate Coordinator:

Professor Salvatore Cafiso – University of Catania

Catania - Toronto, 2013

SUMMARY

INTRODUCTION

The targets of the thesis are road safety analysis based on crash event and road features for the benefit cost analysis where a treatment has to be applied.

There is a growing attention to road safety in Europe. New regulations applied on TERN network push Agencies to introduce new methodological approaches to Road Safety, monitoring the treatment and controlling the level of safety on the managed road network.

A crash is defined as a set of events that result in injury or fatality, due to the collision involving one motorized vehicle or a motor vehicle and another motorized vehicle, a bicyclist, a pedestrian or an object. The terms “crash”, “collision” and “accident” are typically used interchangeably.

“Crash frequency” is defined as the number of crashes occurring at a particular site, in a reference time period.

“Crash rate” is the number of crashes that occurs at a given site during a certain time period in relation to a particular measure of exposure (e.g., per million vehicle miles of travel for a roadway segment or per million entering vehicles for an intersection). Crash rates may be interpreted as the probability (based on past events) of being involved in an accident per unit of the exposure measure.

Accidents count observed at a site (road segment, intersection, interchange) is commonly used as a fundamental indicator of safety performing road safety analysis methods.

Because crashes are random events, crash frequencies naturally fluctuate over time at any given site. The randomness of accident occurrence indicates that short term crash frequencies alone are not a reliable estimator of long-term crash frequency. If a three-year period of crashes were used as the sample to estimate crash frequency, it would be difficult to know if this three-year period represents a high, average, or low crash frequency at the site.

This year-to-year variability in crash frequencies adversely affects crash estimation based on crash data collected over short periods. The short-term average crash frequency may vary significantly from the long-term average crash frequency. This effect is magnified at study locations with low crash frequencies where changes due to variability in crash frequencies represent an even larger fluctuation relative to the expected average crash frequency.

The crash fluctuation over time makes it difficult to determine whether changes in the observed crash frequency are due to changes in site conditions or are due to natural fluctuations. When a period with a comparatively high crash frequency is observed, it is statistically probable that the following period will be followed by a comparatively low crash frequency. This tendency is known as regression-to-the-mean (RTM), and also is evident when a low crash frequency period is followed by a higher crash frequency period.

Failure to account for the effects of RTM introduces the potential for “RTM bias”, also known as “selection bias”. Selection bias occurs when sites are selected for treatment based on short-term trends in observed crash frequency.

RTM bias can also result in the overestimation of the effectiveness of a treatment (i.e., the change in expected average crash frequency). Without accounting for RTM bias, it is not possible to know if an observed reduction in crashes is due to the treatment or if it would have occurred without the modification.

Another key analytical issue arises when accident rates are used in evaluating safety performance to, e.g., flag locations for safety investigation during the network screening process. AADTs are used directly in the computation of this measure, i.e., accident rate = accident frequency/AADT (or some scalar multiple of this). If accident rates are based on the observed counts, then the regression-to-the-mean difficulty discussed above will still apply. In addition, there is an additional problem that renders this method of screening dubious. The problem is that the relationship between accident frequency and AADT is not linear.

Specifically, comparing accident rates of two entities at different traffic levels to judge their relative safety may lead to erroneous conclusions and saying that when two rates are equal they indicate equivalent levels of hazard may be completely false if different AADT levels are involved.

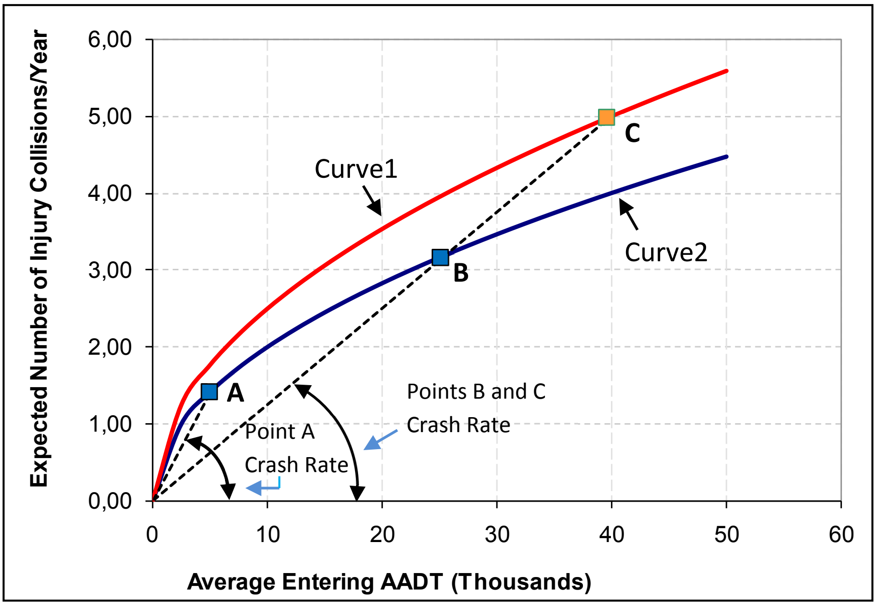


Figure I.1 – Relationship between crash frequency and AADT

See Figure I.1 where two different curves (Curve1 and Curve2) are plotted to give an example. The two curves show the non linear relationship between entering AADT and the expected crash frequency, potentially for two different groups of intersections. It is clear that, in terms of crash rate (the slopes corresponding to each point of the curves), considering accident rates of points A and B, as well as A and C, to judge their relative safety, the comparison may lead to erroneous conclusions because points B or C could be considered by far safer than point A. Moreover, considering points B and C, the two corresponding rates are equal indicating equivalent levels of hazard, but this appears to be completely false because different AADT levels are involved.

The upshot of all this is that the use of accident rate to compare sites in regard to their safety levels is potentially problematic. When the slope of the accidents/AADT relationship is decreasing with increasing traffic volume levels, as is often the case, network screening by accident rates will tend to identify low AADT sites for further investigation. The most valid basis of comparison using accident rates is for the relatively rare cases when the traffic volume levels are the same or when the relationship between accidents and AADT is linear.

Conventional procedures for identifying sites for safety investigation tend to select sites with high accident counts and/or accident rates. However, accident counts could be high or low in a given period solely due to random fluctuations, leading to many sites either incorrectly identified or overlooked and, correspondingly, to an inefficient allocation of safety improvement resources. In addition, selection on the basis of accident rates tends to wrongly identify sites with low volumes. Empirical Bayes approaches have been proposed of late to overcome these difficulties.

The Highway Safety Manual (HSM) in its first edition (2010) introduced a new common approach to the modeling of crash analysis with the aims to standardize the methodology.

In its Part C the HSM provides a predictive method for estimating expected average crash frequency (including by crash severity and collision types) of a network, facility, or individual site. The estimate can be made for existing conditions, alternatives to existing conditions (e.g., proposed upgrades or treatments), or proposed new roadways. The predictive method is applied to a given time period, traffic volume, and constant geometric design characteristics of the roadway. The predictive method provides a quantitative measure of expected average crash frequency under both existing conditions and conditions which have not yet occurred. This allows proposed roadway conditions to be quantitatively assessed along with other considerations such as community needs, capacity, delay, cost, right-of-way, and environmental considerations. The predictive method can be used for evaluating and comparing the expected average crash frequency of situations like:

- Existing facilities under past or future traffic volumes;
- Alternative designs for an existing facility under past or future traffic volumes;
- Designs for a new facility under future (forecast) traffic volumes;
- The estimated effectiveness of countermeasures after a period of implementation;

- The estimated effectiveness of proposed countermeasures on an existing facility (prior to implementation).

Each chapter in Part C of HSM provides the detailed steps of the predictive method and the predictive models required to estimate the expected average crash frequency for a specific facility type like:

- Rural Two-Lane Two-Way Roads
- Rural Multilane Highways
- Urban and Suburban Arterials

However the application of the HSM doesn't always provide adequate results in Europe (Cafiso et al. in 2012 applied the HSM methodology using data of a motorways in Italy in a study published on Procedia Elsevier - Social and Behavioral Sciences, titled "Application of Highway Safety Manual to Italian Divided Multilane Highways", and Sacchi et al. in 2011 studied the transferability of the HSM models for intersection in Italy in a paper published on TRB titled "Assessing international transferability of the Highway Safety Manual crash prediction algorithm and its components"). The problem related to transferability of the Safety Performance Functions (SPFs) is a clear example of how the model developed in other Countries are not always able to catch the safety level of different infrastructures. The key point is that quantification of the expected reduction of crashes related to different treatments, can affect choices and plays a fundamental role in the decision making process. The transferability in HSM is assessed using a factor of calibration to local condition. In this way using equation calibrated on "standard condition" and reported on the Manual, the application of the models in the whole North America is assessed. However many Authors have tried to apply HSM in Europe with poor results.

Together with SPFs the Highway Safety Manual introduced the "Crash Modification Factor" (CMF). A Crash Modification Factor is a multiplicative factor used to compute the expected number of crashes after implementing a given countermeasure at a specific site. The CMF is multiplied by the expected crash frequency without treatment. A CMF greater than 1.0 indicates an expected increase in crashes, while a value less than 1.0 indicates an expected reduction in crashes after implementation of a given countermeasure.

The best methodology of estimation of CMFs is well known to be based on stochastic approach. The problem of regression to the mean and the selection bias can be controlled using a sophisticated probabilistic approach reported in the "Observational Before/After Studies in Road Safety - Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety" by Hauer in 1997 and developed by various author in the last 2 decades.

The new methodologies developed for the calibration of the Safety Performance Functions are pushing the Authors (See Persaud et al. "Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations" Published on Accident Analysis and Prevention in 2009) to find new advanced methodology able to address the problem of time trend and spatial correlation of data and to use more complicated model form and different distribution of outcomes in the calibration of CMFs. Despite the efforts on the calibration of CMFs to improve reliability, evaluation of safety benefits of applying a treatment continue to be performed using a deterministic approach. However in the HSM there is not a methodology to assess the CMFs transferability, and their stochastic nature doesn't allow a perfect reliability also if they are applied in the same Country. Some Authors are developing different methodology to assess the transferability of the CMFs, and they should be applied when a CMF has to be used in a benefit cost analysis. The traditional techniques for the evaluation of the benefits of a treatment don't take into account the statistical distribution of the CMFs and their stochastic nature.

Objectives of this Thesis and Overview of the Contents

The regression to the mean effect as well as the functional relationship between crashes and exposure factors don't allow a reliable estimation of the effect of a treatment. Have a high reliability in the estimation of the effects of a countermeasures may be one of the most important issue in a Benefit-Cost analysis. The identification of hazardous location together with the evaluation of the alternatives to fix safety problem, are based on the reliability of the model used for the evaluation. For those reason the research work provide a wide discussion on the main step for a reliable benefit-cost analysis starting from different methodology able to address the problem of regression to the mean and time trend effects and able to optimize the goodness of fit of the models varying the segmentation approach and the variables considered in the calibration of the models. The second step is to consider a new methodology for the evaluation of the variance of the CMFs. Particularly if a cross site variance is considered together with the variance of the CMF, as an indicator of the distribution of the CMF in different site, the perspective of considering the effect of a treatment can change drastically. The main objective of the present research work is to develop a methodology to perform Benefit-Cost analysis taking into account a cross site variability of the CMFs and their variance.

To do that an overview on the modeling approach is reported as well as a calibration of a Crash Modification Factor for safety barrier in Italy. At the end the proposed methodology for the stochastic benefit cost analysis is detailed described in comparison with the deterministic approach.

In the first Chapter an overview about SPFs is reported, with a wide literature review on the topic. In the second Chapter a methodology to address the time trend effects in the calibration of SPFs is applied using data on a motorways in Italy, the A18 Messina - Catania. Particularly the second Chapter focus on the problem of time trend, generally presents when a long period of analysis is taken into account.

The third Chapter focuses on one of the possible method to optimize the model goodness of fit when roadway segment are analyzed. Particularly the segmentation approach was investigated and tested with the more common goodness of fit evaluation method. At the end of the Chapter a ranking was performed comparing the EB methodology to observed number of crashes and a methodology for the identification of hazardous location was applied.

In the fourth Chapter various methodologies to estimate a CMF are reported. As a case study an application of the empirical Babyes methodology is reported for the estimation of a CMF for road safety barrier. In the second part of the Chapter a Function was calibrated considering the functional relationship between the barrier direct related categories of crashes (ran-off-road crashes) and curvature using the same data of the presented case study.

The finals 5th and 6th Chapters deal with the benefit-cost analysis. The cross site variance was evaluated for the CMF calibrated in Chapter 4 and used in the stochastic benefit-cost analysis. A comparison on the traditional approach and a stochastic approach was performed for new and existing infrastructures. At the end of the Chapter a methodology to combine more CMFs in the same segment was described.

CHAPTER 1

SAFETY PERFORMANCE FUNCTIONs (SPFs): A GENERAL OVERVIEW

1.1. Introduction

Road crash events are the object of many studies because of their potentially severe consequences. Thus, it is not a surprising that particular attention is given to the development of models able to identify features related to accidents and to forecast accident frequency.

Many researchers have developed accident prediction models also known as safety performance functions (SPFs) calibration. The high number of factors and the correspondingly high combinations of these factors related to accident event occurrence created need to develop different models for different circumstances. Models were developed specifically for urban and rural roads, for road segments and intersections; models account for different types of roads and intersections or even for different accident types and severity. Moreover, for the same safety performance function, the variables considered significantly related to safety, as well as the statistical regression approaches adopted can vary considerably.

The statistical methodologies mainly adopted for calibration of safety performance functions, are the conventional linear regression and the generalized linear regression techniques. Since the mid-eighties, many research works highlighted the limitations of the conventional linear regression approach, which is why, nowadays the generalized linear regression technique (GLM) is the most common statistical regression methodology used. Nevertheless in the last two decades new techniques of calibration were carried out. One is the full Bayes methodology of calibration, which is able to account for the effect of regression to the mean and the Generalized Estimating Equation (GEE) able to account for the time correlation more extensive described in the Chapter 2. In the Present Chapter 1, a wide description of the model approach is reported with a literature review and the key points of the GLM methodology are explored, with particular attention to the reliability of the models.

1.2. <u>Literature overview</u>

Following, in chronological order, is a wide literature review regarding the statistical approaches adopted and results achieved by various studies and researchers. The literature overview starts from the early nineties and discusses in more depth the latest works dealing with the generalized linear modeling statistical regression technique. In general, the reviews merely summarize the papers, including not only the approach and findings but also verbatim expressions of the authors' opinions and beliefs and their own review of related work.

Review of Persaud and Dzbik (1993) [1.1]

These authors proposed a freeway accident modeling approach with the aim to overcome limitations of previous models.

In the authors' opinion, the first difficulty with existing models is that they tend to be macroscopic in nature since they relate accident occurrence to average daily traffic (ADT) rather than to the specific flow at the time of accidents. Second, some modelers assume, a priori, that accidents are proportional to traffic volume and go on to use accident rate (accident per unit of traffic) as the dependent variable. There is much research to suggest that this assumption is not only incorrect but can also lead to paradoxical conclusions [1.2]. Third, conventional regression modeling assumes that the dependent variable has a normal error structure. For accident counts, which are discrete and nonnegative, this is clearly not the case; in fact, a negative binomial error structure has been shown to be more appropriate [1.3]. Finally, it is impossible for regression models to account for all of the factors that affect accident occurrence.

The need to overcome these difficulties was fundamental to the modeling approach adopted in the authors' work. To this end, the authors adopted a generalized linear modeling statistical approach that allows the flexibility of a nonlinear accident-traffic relationship and the possibility to choose a more appropriate error-structure for the dependent variable. The approach was applied to both microscopic data (hourly accidents and hourly traffic) and macroscopic data (yearly accident data and average daily traffic).

Generalized linear modeling using the GLM computer package was used to obtain a regression model for estimating P, the accident potential per kilometer per unit of time, given a freeway section's physical characteristics, the volume (T) per unit of time, and a set of variables that describe operating conditions during the time period. The model form used was:

$$E(P) = a \cdot T^{b} \tag{1.1}$$

where a and b are model parameters estimated by GLM.

The macroscopic models were calibrated using data obtained from the Ontario Ministry of Transportation. For approximately 500 freeway sections, the accident count for the years 1988 and 1989 were used as an estimate of the dependant variable, and traffic data as an independent variable. To account for varying section lengths, the term log (section length) was specified as an "offset", thus, in effect, models were estimated for prediction of the number of accidents per kilometer per year.

The microscopic models were calibrated using data pertaining to a 25-km segment of Highway 401 in Toronto, Canada. Part of this freeway, actually, has a Traffic Management System (FTMS) which provided detailed traffic data for short time periods. The sections, which range in length from 0.7 to 3.0 km, are separated by interchanges, and all have express and collector roadways typically with three lanes each per direction.

It was decided to disaggregate each day into 24 periods of 1 hour each and to derive data for each hour, for express and collector lane separately, and for day and night. For the accident data this task was straightforward. For the traffic data, it was necessary to derive hourly and seasonal variation factors and collector/express lane distribution factors and apply these factors to the average daily traffic. To maintain a reasonable level of homogeneity, only data pertaining to weekdays were used for the model calibration. After preliminary data analysis, it was decided to build the regression models using, for each section, only data for off-peak hours for which that section tended to be uncongested.

Figure 1.1 shows plots of the microscopic model regression prediction per kilometer per hour for two accident types (severe and total) and for express and collector roadways.

It is important to note that, for these regression lines, the slope is decreasing as hourly volume increases, perhaps capturing the influence of decreasing speed.

This is in contrast to the macroscopic plot in Figure 1.2, which all show increasing slopes. It is possible that the macroscopic plots are reflecting the increasing probability of risky maneuvers, such as passing and changing lanes, with higher ADT levels.

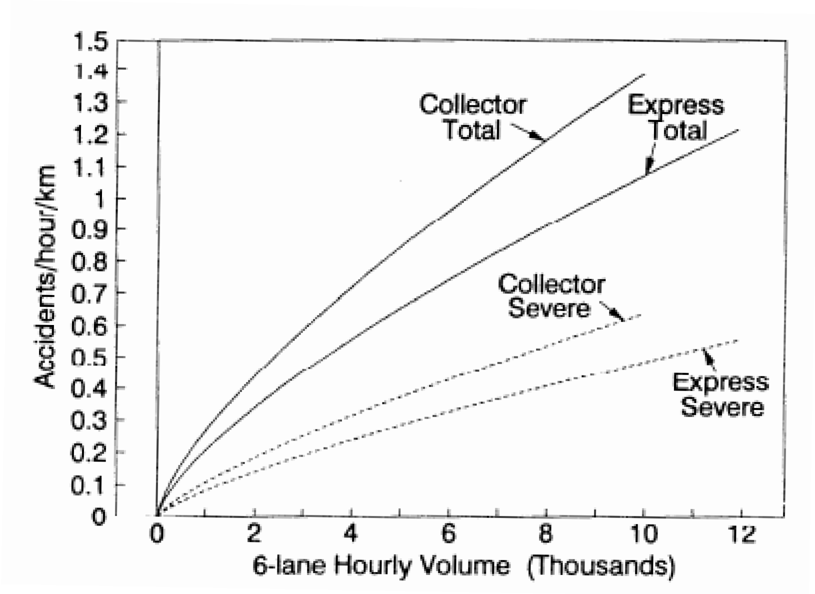


Figure 1.1. Microscopic regression model prediction

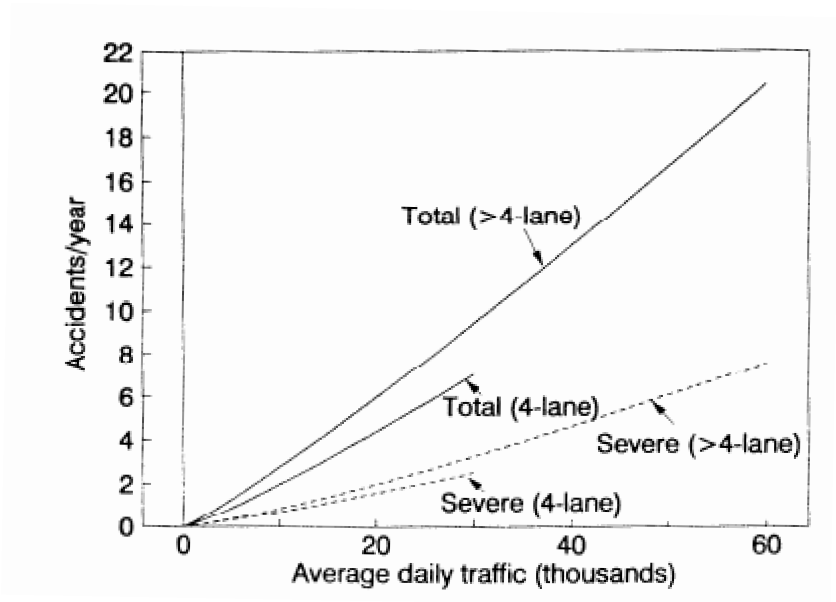


Figure 1.2. Macroscopic regression model prediction

Review of Mountain, Fawaz and Jarrett (1996) [1.4]

These researchers developed and validated a method for predicting expected accidents on main roads with minor junctions where traffic counts on the minor approaches are not available.

The authors recognize that accidents at junctions are ideally modeled separately and that junction models do not normally assume a linear relationship between accidents and conflicting flows [1.5][1.6][1.7][1.8][1.9]. However, problems can arise because the application of such models requires, at a minimum, a knowledge of entry flows.

The data used for the study comprised details of highway characteristics, accidents and traffic flows in networks of main roads in seven UK counties for periods between 5 and 15 years. The networks represented a total of some 3800 km of highway. The road networks were restricted to UK A and B roads outside major conurbations. In addition to road type, the network roads were categorized according to carriageway type (single or dual) and speed limit (<40 mph (urban) and 40> mph (rural)). Junction accidents were defined as accidents occurring within 20 m of the extended kerblines of the junction.

The regression models were developed using the generalized linear modeling technique. This approach allowed a model form in which accidents are not linearly proportional related to traffic and the assumption of a negative binomial error structure for the dependent variable.

Results were compared with the COBA model (a basic model used in the U.K. for use in the cost benefit analysis), in which dependent variable is assumed to be proportional to the link length and traffic flow. It was clear that the proportional model used in COBA increasingly tended to overestimate the annual accidents as traffic volume increase. The non-linear model form showed that the linear model, as used in COBA, is inappropriate.

Review of Abdel-Aty and Radwan (2000) [1.10]

These authors argued that two major factors usually play an important role in traffic accident occurrence. The first is related to the driver, and the second is related to the roadway design.

Actually, as Abdel Aty et al. note, different researchers have attempted three approaches to relate accidents to geometric characteristics and traffic related explanatory variables: Multiple Linear regression, Poisson regression and Negative Binomial regression. However, recent research shows that multiple linear regression suffers some undesirable statistical properties when applied to accident analysis. To overcome the problems associated with multiple linear regression models, researchers proposed Poisson regression for modeling accident frequencies. They argued that Poisson regression is a superior alternative to conventional linear regression for applications related to highway safety. In addition, it could be used with generally smaller sample sizes than linear regression. Using the Poisson model necessitates that the mean and variance of the accident frequency variable (the dependent variable) be equal. In most accident data, the variance of the accident frequency exceeds the mean and, in such case, the data would be over dispersed. Because of the over dispersion difficulties, the authors suggested the use of a more general probability distribution such as the Negative Binomial.

The primary objective of the Abdel-Aty and Radwan research was to develop a mathematical model that explains the relationship between the frequency of accidents and highway geometric and traffic characteristics. Other objectives include developing models of accident involvement for different gender and age groups using the Negative Binomial regression technique, based on previous research that showed significant differences in accident involvement between different gender and age groups [1.11][1.12][1.13].

In order to develop a mathematical model that correlates accident frequencies to the roadway geometric and traffic characteristics, Abdel Aty et al. argue that one needs to select a roadway that possess a wide variety of geometric and traffic characteristics. The goal of their data collection exercise was to divide the selected roadways into segments with homogenous characteristics. After reviewing several roadways in Central Florida, the authors decided that State Road 50 (SR 50) was most appropriate for this task. Information included geometric characteristics such as horizontal curves, shoulder widths, median widths, and traffic characteristics such as traffic volumes and speed limits. SR 50 was divided into 566 highway segments defined by any change in the geometric and/or roadway variables. The data included the following variables: AADT, degree of horizontal curvature, shoulder type, divided/undivided, rural/urban classification, posted speed limit, number of lanes, road surface and shoulder types, and lane, median, and shoulder widths. Accident data were obtained from an electronic accident database for three years from 1992 to 1994.

The Poisson regression methodology was initially attempted. However, the Poisson distribution was rejected because the mean and variance of the dependent variables are different, indicating substantial over dispersion in the data. Such over dispersion suggested a Negative Binomial model to the authors who note that the Negative Binomial modeling approach is an extension of the Poisson regression methodology and allows the variance of the process to differ from the mean. The Negative Binomial model arises from the Poisson model by specifying:

$$ln\lambda_i = \beta x_i + \varepsilon \quad (1.2)$$

Where, λ_i is the expected mean number of accidents on highway section i; β is the vector representing parameters to be estimated; x_i is the vector representing the explanatory variables on highway segment i; ε is the error term, where exp(ε) has a gamma distribution with mean 1 and variance α^2. The Negative Binomial model is calibrated by standard maximum likelihood methods. The likelihood function is maximized to obtain coefficient estimates for α and β. The choice between the Negative Binomial model and the Poisson model can largely be determined by the statistical significance of the estimated coefficient α. If α is not significantly different from zero (as measured by t-statistics) the Negative Binomial model simply reduces to a Poisson regression.

In order to decide which subset of independent variables should be included in an accident estimation model, AIC (Akaike's information criterion) was used by Abdel Aty et al. AIC identifies the best approximating model among a class of competing models with different numbers of parameters. AIC is defined as $AIC=-2 \cdot ML+2 \cdot p$, where ML is the maximum value of the log-likelihood function and p is the number of the variables in the model. The smaller the value of AIC, the better the model.

Two exposure variables were found to be significant. The first is the section's length. The longer the length of the roadway section, the more likely accidents would occur on these sections. A similar conclusion was reached for the log of the AADT per lane. Moreover, the sharpness of the horizontal curve has a positive effect on the likelihood of accidents. Accidents increase with increasing degree of curve. An increase in shoulder width and median width was associated with a reduced frequency of accidents. There was an interaction effect between the lane width and the number of lanes. When the lane width increases, and at the same time the number of lanes decreases, the frequency of accidents decline. Vertical alignment did not enter the model, possibly because Florida has relatively flat topography (i.e. little variation in slopes). Finally, the significance of the over dispersion parameter (α) indicates that the Negative Binomial formulation is preferred to the more restrictive Poisson formulation.

From the male and female accident involvement models, it could be concluded that female drivers experience higher probability of accidents than male drivers during heavy traffic volume and with reduced median width. Moreover, narrow lane width and larger number of lanes have more effect on accident involvement for female drivers than male drivers. Male drivers have greater tendency to be involved in accidents while speeding.

Young and older drivers have a larger possibility of accident involvement than middle aged drivers when experiencing heavy traffic volume. There is no effect of horizontal curve on older drivers' accident involvement. Older age drivers, however, have a greater tendency to accident occurrence than middle and young drivers for reduced shoulder width and median widths. Decreasing lane width and increasing number of lanes creates more problems for older drivers and younger drivers than middle age drivers. Older drivers experience fewer accidents if the shoulder is paved. Also, the likelihood of younger drivers' accident involvement increases with speeding.

Review of Garber and Ehrhart (2000) [1.14]

This study identified the speed, flow, and geometric characteristics that significantly affect crash rate and developed mathematical relationships to describe the combined effect of these factors on the crash rate for two-lane highways.

The authors noted that previous research of crash rates and hourly traffic volume revealed a U-shaped relationship, indicating that higher crash rates are observed during the early-morning and late-day hours when the traffic volume is low [1.15][1.16]. This has been further extrapolated to conclude that, as the number of vehicles on the highway increases, the variation between vehicle speeds decreases and that it is the speed variance that affects the crash rate [1.17][1.18][1.19]. The type of crash also has been shown to be a function of the traffic volume. The percentage of multiple-vehicle crashes decreases as the traffic volume decreases, and the percentage of single-vehicle crashes increases with a decrease in volume [1.20].

Garber et al. cited research showing that a U-shaped relationship exists between the probability of a vehicle being involved in a crash and the deviation of the vehicle's speed from the mean speed of the traffic. This relationship indicates that the greater a vehicle speed deviates from the mean speed, the greater is the probability of that vehicle being involved in a crash [1.20]. This implies that driving both slower and faster than the mean speed increases the likelihood of being involved in a crash, according to Garber et al.

The review by Garber et al. found that the main geometric characteristics that have been found to influence safety for two-lane roads are lane width and shoulder width. The results obtained from these studies have tended to be inconsistent. A few studies also have investigated the effect of grade on crashes for two-lane highways. Although it is generally accepted that grades of 4 percent or lower have an insignificant effect on crashes, the results of these studies also have been contradictory.

In the actual Garber and Ehrhart (2000) research, data collection consisted of obtaining speed, flow, and crash data for the road segments selected for the study and defining roadway characteristics, such as the number of lanes, for the segments. Lengths of the roadway segments were identified between traffic monitoring stations, positioned on major intersections, to ensure homogeneous traffic and flow characteristics.

The modeling process began with the use of the independent variables mean speed (MEAN), standard deviation of speed (SD), flow per lane (FPL), lane width (LW), and shoulder width (SW), and with crash rate (CRASHRATE) as the dependent variable. Two deterministic modeling procedures (multiple linear regression and multivariate ratio of polynomials) were applied to the data in search of an adequate fit, but in the end, multivariate ratio of polynomial models seemed to better described the relationship between the dependent and independent variables.

The authors recognized that once models have been constructed, a suitable measurement of quality must be applied to select the model that best fits the observed data. Two such measurements were applied in this research: the coefficient of determination (R^2) and Akaike's information criterion (AIC).

The R^2 value is a popular measure used to judge the adequacy of a regression model. Defined as a ratio, the R^2 value is a proportion that represents the variability of the dependent variable that is explained by the model. In symbolic form:

$$R^2 = \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \tag{1.3}$$

where:

$\hat{y}$: model estimates,
$\bar{y}$: mean of the observations
y_i: actual observations

An R^2 value near zero indicates that there is no linear relationship between the dependent and independent variables, while a value near 1 indicates a linear fit. The authors note that the R^2 value should be used with caution to ensure its correct interpretation, and it always should be accompanied by an examination of the residual scatter plots; also, that the R^2 value is not an appropriate measure for nonlinear regression because its purpose is to measure the strength of the linear component of the model. Therefore, another method must be introduced to compare the multivariate models. The method chosen was the AIC, which was developed as a means to predict the fit of a model based on the expected log likelihood.

The results of this research established that the crash rate is not linearly related to speed, flow, and geometric characteristics. Of all of the independent variables considered in this study, the standard deviation of speed (SD) seemed to have the greatest impact on the crash rates for two-lane highways.

Review of Pardillo and Llamas (2003) [1.21]

These researchers developed a set of multivariate regression models to estimate crash rates for two-lane rural roads, using information on accident experience, traffic and infrastructure characteristics. The research includes identifying relevant variables, model calibration and precision analysis of the method for accident rate prediction and for assessment of road safety improvement projects effectiveness. In the development of accident prediction models two key questions have to be solved, as the authors note: functional model form choice and independent variable selection.

The authors felt that in recent years, there was a consensus among researchers in favor of modelling accidents as discrete, rare, independent events, usually as generalized linear Poisson models in which the frequency of crashes that occurs in a given road section is treated as a random variable that takes discrete integer non-negative values. A characteristic feature of the distribution is that the variance of the variable is equal to its mean. The mean number of accidents is assumed to be an exponential applied to a suitable linear combination of road variables. The resulting models are generalized linear models (GLIM), in which the exponential function guarantees that the mean is positive. The authors cited Maher and Summersgill [1.22] who developed a comprehensive methodology to fit predictive accident models applying this approach.

In 1994 Miaou [1.23] introduced negative binomial models, a generalization of the Poisson form that allows the variance to be over-dispersed and equal to the mean plus a quadratic term in this mean whose coefficient is called the overdispersion parameter. When this parameter is zero, a Poisson model results.

Vogts and Bared [1.24] developed a series of Poisson and Negative Binomial multivariate regression models to predict accident frequencies in 2-lane rural roads and intersections. Prediction variables in non-intersection models include traffic volume, commercial vehicles percentage, lane and shoulder width, horizontal and vertical alignment, road side condition and driveway density.

Independent variable selection for accident prediction models remains a complicated problem. Krammes et al. [1.25] and Lamm et al. [1.26] have both shown with their works the importance of taking into account design consistency when considering safety effects of highway characteristics on crash risk.

A sample 3450 km of two-lane rural roads was used in this research. Crash data was obtained and analyzed in two periods: 1993-97 and 1998-99. The first period was used in model calibrations, while the second period was reserved to assess model accuracy. Traffic and roadway characteristics dataset contains one record every 10 m of roadway including the following data:

- AADT (veh/day)
- Curvature(m^{-1})
- Longitudinal grade (%)
- Roadway width (m)
- Right shoulder width (m)
- Left shoulder width (m)
- Sight distance (m)
- Access points
- Posted speed limit (km/h)
- No passing zones
- Safety barriers

Dividing the sample in homogeneous sections in which all the characteristics of the highway were constant resulted in segments with of an average length of less than 400 m. Previously, in 1997, Resende and Benekohal [1.27] had reached the conclusion that to get reliable accident prediction models crash rates should be computed from 0.8 km or longer sections. It was decided that the average length of homogeneous sections was too short to allow for a meaningful analysis of the effect of potential roadway improvements on safety. For that reason, the study was conducted in parallel for two types of sections. For the first one the 3450 km sample was divided in 1 km fixed length segments. For the second, the same sample was divided into 236 highway sectors or sections of variable length limited by major intersections and/or built-up areas within which traffic volumes and characteristics could be assumed to be constant. The length of these highway sectors ranged between 3 km and 25 km, with an average of 14.6 km.

The variables that were considered in the analysis for the 1 km long segments were:

- Access density (access points/km)
- Average roadway width (m)
- Minimum sight distance (m)
- Minimum curvature (1/m)
- Minimum speed limit (km/h)
- Maximum grade in absolute value (%)
- Minimum design speed (km/h)
- Design speed reduction from the adjacent 1 km segments (km/h)

The variables that were considered in the analysis for the highway sectors or sections of variable length were:

- Access density (access points /km)
- Average roadway width (m)
- Average sight distance (m)
- Average curvature (1/m)

- Standard deviation of curvature (1/m)
- Average speed limit (km/h)
- Maximum longitudinal grade (%)
- Average of the absolute values of grade (%)
- Average design speed (km/h)
- Standard deviation of design speed values (km/h)
- Average design speed variation between the 1 km long adjacent segments included in the sector (km/h)
- Proportion of no passing zones

Independent variable selection was performed with the objective of identifying those variables that show higher degree of association with crash rates. Traffic volume was found to be the variable with the highest correlation with crash frequencies.

In 1 km segments, the highest correlation coefficients with the average accident rate for the study period were:

- Access density (access points/km)
- Design speed reduction from adjacent segments (km/h)
- Speed limit (km/h)
- Average sight distance

When longer sections were considered, the highest correlation coefficients with the average accident rate were:

- Access density (access points/km)
- Average speed limit (km/h)
- Average sight distance (m)
- Proportion of no-passing zones

From the results of the research performed by Pardillo and Llamas, the following conclusions were obtained:

- A key step to develop accident prediction models is to select a set of independent variables that capture as much of the interaction between roadway characteristics and driver safety performance as possible. To do this a univariate correlation analysis can be conducted prior to the calibration of multivariate models.
- The highway variables that have the highest correlation with crash rates in Spain´s two-lane rural roads are: Access density, average sight distance, average speed limit and the proportion of no-passing zones. Access density is the variable that influences most the rate of head-on and lateral collisions, while in run-off the road and single vehicle crashes sight distance is decisive.
- High access density has a negative effect on safety. Therefore preventive safety improvements should include access management and control measures.
- To measure the influence of geometric design on crash rates it is necessary to use variables that measure the variation of characteristics between adjacent alignment elements or along a highway section. This confirms the importance of achieving highway design consistency to improve safety.

Review of Ng and Sayed (2004) [1.28]

The objectives of this study were to investigate and quantify the relationship between design consistency and road safety.

Design consistency is the conformance of geometry of a highway with driver expectancy, and its importance and significant contribution to road safety is justified by understanding the driver–vehicle–roadway interaction. When an inconsistency exists that violates driver's expectation, the driver may adopt an inappropriate speed or inappropriate maneuver, potentially leading to accidents. In contrast, when design consistency is ensured, all abrupt changes in geometric features for continuous highway elements are eliminated, preventing critical driving maneuvers and minimizing accident risk (Fitzpatrick and Collins 2000) [1.29].

Currently, several measures of design consistency have been identified in the literature and models have been developed to estimate these measures. These measures can be classified into four main categories: operating speed, vehicle stability, alignment indices, and driver workload.

Operating speed is a common and simple measure of design consistency. Operating speed is defined as the speed selected by the drivers when not restricted by other users, i.e., under free flow conditions, and it is normally represented by the 85th percentile speed, denoted as V_{85} (Poe et al. 1996) [1.30]. The difference between operating speed and design speed (V_{85}-V_d) is a good indicator of the inconsistency at one single element, while the speed reduction between two successive elements (ΔV_{85}) indicates the inconsistency experienced by drivers when traveling from one element to the next. Lamm et al. (1999) [1.31] have established consistency evaluation criteria for these measures.

Many models have been developed to determine operating speed in terms of alignment parameters. Morrall and Talarico (1994) [1.32] have related V_{85} (kilometers per hour) on horizontal curves to the degree of curve DC using data on two-lane rural highways in Alberta.

Lamm et al. (1999) have suggested that another measure that can account for more variability of operating speed on curves is the curvature change rate CCRs because it takes transition curves into consideration, as shown below

$$\mathrm{CCR_s} = \frac{63700(L_{\mathrm{cl1}}/2R + L_{\mathrm{cr}}/R + L_{\mathrm{cl2}}/2R)}{L} \tag{1.4}$$

where CCR_s is measured in gon per kilometers (gon is a designation of the angular unit (1 gon = 0.9°)), L_{cr} is the length of circular curve (meters), L_{cl1} and L_{cl2} are the lengths of spirals preceding and succeeding the circular curve (meters), and $L = L_{cr} + L_{cl1} + L_{cl2}$ is the total length of curve and spirals (meters).

Operating speed on tangents connecting horizontal curves of an alignment is also important for design consistency evaluation. Tangent length is one of the factors that determines the necessary speed reduction when entering a horizontal curve and forms the basis of the definitions of dependent and independent tangents. An independent tangent is a tangent that is long enough to allow drivers to reach their desired operating speed, and consequently a speed reduction of greater than 20 km/h is required when they enter the following curve (Lamm et al. 1999). In contrast, a nonindependent tangent is one that is not long enough, and therefore a speed reduction that is greater than 20 km/h is not required. Speed on nonindependent tangents is less complex to model than that on independent tangents that generally depend on a whole array of roadway character. Some research has been undertaken to model operating speed on independent tangents, but the results are considered preliminary (Polus et al. 2000) [1.33].

Alignment indices are defined as quantitative measures of the general character of an alignment. They reveal geometric inconsistencies when the general characteristics of the alignment change significantly. While speed reduction and vehicle stability are good measures of design consistency, they are symptoms rather than causes. It is the geometric design itself, or specifically, the geometric characteristics and the combinations of tangents and horizontal curves that create inconsistencies. One of the indicators of geometric inconsistency is a large difference between the value of an alignment index of an individual feature and the average value of the alignment. The ratio of the radius of an individual horizontal curve to the average radius of the alignment (roadway section), denoted by CRR, is adopted in Ng and Sayed research work.

Driver workload can be defined as the time rate at which drivers must perform the driving task that changes continuously until it is completed (Messer 1980) [1.34]. Conceptually, driver workload can be a more appealing approach for identifying inconsistencies than operating speed because it represents the demands placed on the driver by the roadway, while operating speed is only one of the observable outputs of the driving task. However, the use of driver workload is much more limited than operating speed because of its subjective nature (Krammes and Glascock 1992) [1.35].

Two measures have been proposed in the literature to measure driver workload, namely sight distance and visual demand. Limited sight distance increases driver workload as the driver needs to update his information more frequently and process it more quickly. However, little research has been conducted to investigate the relationship between driver workload and sight distance. In contrast, a number of studies have been carried out to examine the potential of visual demand as a measure of driver workload. Visual demand is defined as the amount of visual information needed by the driver to maintain an acceptable path on the roadway (Wooldridge et al. 2000) [1.36].

The models developed to estimate visual demand of unfamiliar drivers (VD_{LU}) and visual demand of familiar drivers

(VD_{LF}) were found to be inversely proportional to horizontal radius, meaning that driver workload increases with a decrease in radius [1.37].

The study conducted by Ng and Sayed uses geometric design, accident, and traffic volume data recorded on a two-lane rural highway. Specifically, accidents that occurred within 50 m of signalized intersections or within 20 m of all other types of intersections were eliminated. The design consistency measures mentioned previously, namely V_{85}-V_d, ΔV_{85}, CRR, VD_{LU}, and VD_{LF}, were computed for each section. The data included 319 horizontal curves and 511 tangents.

The methodology used in this study is based on the development of Accident Prediction Models incorporating design consistency measures. The models are developed using the generalized linear regression modeling (GLM) approach. The following model form was adopted:

$$E(\Lambda) = a_0 L^{a_1} V^{a_2} e^{\sum_{j=1}^{m} B_j x_j} \qquad (1.5)$$

where $E(\Lambda)$ is the expected accident frequency; L is the section length; V is the annual average daily traffic (AADT); x_j is any of the m variables in addition to L and V; and a_0, a_1, a_2, and b_j are model parameters. The error structure of the models is assumed to follow the negative binomial distribution.

Models developed showed accident frequency to be positively correlated to V_{85}-V_d, ΔV_{85}, VD_{LU}, and VD_{LF}, and is negatively correlated CRR.

A qualitative comparison was also made to compare accident prediction models that explicitly consider design consistency with those that rely on geometric design characteristics for predicting accident occurrence. The comparison, while limited to fictitious alignments and not real data, shows that the first type may be superior as it can potentially locate more inconsistencies and reflect the resulting effect on accident potential more accurately than the second. The prediction accuracy of accident prediction models is limited by the quality of their independent variables. As such, the models developed in this study depend heavily on the design consistency measures used.

Review of Zhang and Ivan (2005) [1.38]

These researchers used Negative Binomial (NB) Generalized Linear Models (GLIM) to evaluate the effects of roadway geometric features on the incidence of head-on crashes on two-lane rural roads in Connecticut.

Many previous studies have applied NB GLIM in highway crash analysis under different circumstances. Wang and Nihan (2004) [1.39] used NB GLIM to estimate bicycle-motor vehicle (BMV) crashes at intersections in the Tokyo metropolitan area. Shankar et al. (1995) [1.40] also adopted NB GLIM in modeling the effects of roadway geometric and environmental features on freeway safety. Miaou (1994) [1.41] evaluated the performance of negative binomial regression models in establishing the relationship between truck crash and geometry design of road segments.

For Zhang and Ivan research, six hundred fifty-five segments, each with a uniform length of one kilometer, were selected from fifty Connecticut state-maintained two-lane rural highways. The selection was based on the land use pattern, permitting only minor intersections (without signal or stop control on the major approaches) and driveways along the segments. Information concerning speed limit, clear roadway width, number of driveways and minor intersections, and geometric characteristics such as the horizontal curvature and the vertical grade were collected. The definitions of the selected variables are shown in Table 1.1.

Table 1.1. Definition of Site Characteristic Variables (Zhang and Ivan, 2005)

Direct Variable	Description
SW	Paved shoulder width, ranging from 2 to 8 feet
LW	Paved lane width, typical values (10 feet, 12 feet, 13 feet)
Width	Paved roadway width (SW+LW) (directional), ranging from 10 to 20 feet
AccessP	The number of driveways and minor intersection in each segment
SpeedLT	The speed limit, ranging from 25 to 50 MPH
Surrogate Variable	**Description**
WMAH	Weighted mean of absolute horizontal curvature (horizontal curve radius)
SACRH	Sum of absolute change rate of horizontal curvature
MAXD	Maximum absolute degree of horizontal curvature
WMAV	Weighted mean of absolute vertical curve (grade) (computed as for horizontal curves)
SACRV	Sum of absolute change rate of vertical curve (computed as for horizontal curves)
MINK	Minimum K value of all vertical curves on the segment
CHV	Sum of combined horizontal curvature and vertical curve
LCHV	Logarithmic CHV value

The Akaike's Information Criterion (AIC) [1.42][1.43] was used for selecting the best of the models and to highlight variables significantly related to safety. Four of the site variables were found to be significant at 95 percent confidence for predicting the head-on crash count: SACRH, SACRV, MAXD, and SpeedLT (speed limit). Moreover, the authors found that the coefficient for the natural log of AADT is significantly different from 0, rejecting the hypothesis that the rate of head-on crashes is constant with volume. This coefficient is actually negative, suggesting a decreasing trend for head-on crash rate with AADT. This was not expected, since head-on crashes are expected to occur more often at higher volumes than at lower volumes, as drivers would have more opportunities to conflict with vehicles approaching from the opposite direction. Nevertheless, since head-on collisions are so rare, this relationship may be relatively weak. Also, drivers may pay more attention to safety when they see more traffic coming from the opposite direction, thus reducing the rate of head-on crashes at high traffic volumes.

Consistency of geometric design can be approached in two ways. Some studies have analyzed lengths of road, to address the possibility that accidents may not occur at “the most inconsistent” element (curve or tangent) of the alignment but somewhere within a road section of poor consistency. The more usual approach is to analyze individual elements (curves and tangents) of the alignment. This provides a more specific level of analysis.

Review of Bird and Hashim (2006) [1.44]

A similar approach to Zhang and Ivan was adopted in this research in which authors aimed to find whether relationships could be found between accident locations and various consistency measures at element level. Many consistency variables were defined for this study, some have been used before in other studies, and some were new ones defined by the authors for this particular study. These variables can be alignment indices (ratio of curve radius to average curve radius) or speed indices (difference between operating speed on an element and speed limit).

Other research has been carried out to develop accident prediction models that are based on a range of alignment consistency measures. Anderson (1999) [1.45] developed several regression models to relate the accident frequency to 5 different consistency measures separately. These measures include operating speed reduction, average radius, ratio of an individual radius to average radius and average rate of vertical curvature. His data set consisted of 5287 horizontal curves in 6 states in USA. His final conclusion stated these four consistency measures appeared suitable for assessing the safety of highways. The fifth measure, ratio of maximum to minimum radius on a roadway section, was found not to be as sensitive to predicted accident frequency, and was therefore not recommended as a design consistency measure. Other work on Canadian roads, recently reported by Hassan et al. (2005) [1.46], found that operating speed consistency provided superior models in relation to collision frequency than design-speed margin consistency.

The research proposed by Bird and Hashim (2006), was based on a sample of 380 km of rural single carriageways (two lane undivided highways). Sections in built up areas (villages or towns) or near (within 20 meters of) junctions with other A or B class roads were excluded. The study required geometric details of each element of the alignment of the road, thus alignment characteristics (e.g. length of element, radius, deflection angle and degree of curve) were collected. The final sample contained 620 curves and 594 tangents. The study considers only personal injury accidents (PIA) because in the U. K., only personal injury accidents must, by law, be reported to the police. The time period for the accident data was 2000-2004. Accident records were allocated to the correct elements (e.g. curve or tangent) using the easting and northing coordinates of each accident and element.

One of the variables used in calculating some of the consistency measures is operating (85th percentile) speed. The design speed was also used as part of some consistency indices.

So as already highlighted previously, the technique of Generalized Linear Modelling (GLM) offers a suitable and sound approach for developing accident prediction models. The general form of the accident prediction model is therefore:

$$\mu = \exp(\beta_0) AADT^{\beta_1} L^{\beta_2} \exp(\sum_{j=3}^{n} \beta_j X_j) \tag{1.6}$$

where:

μ = estimated accident frequency

L = section length

AADT = section average annual daily traffic,

X_j = any additional variables

β_0, and β_j = the regression parameters

The usual test for goodness of fit for standard regression analysis is the R^2 value. However this has shortcomings when used in accident analysis as stated by Miaou (1995) [1.47]. Many other measures have been suggested by Miaou and other authors [1.48][1.49][1.50]. Miaou suggested using the negative binomial overdispersion parameter k to determine how well the variance of data is explained in a relative sense. This is expressed as:

$$R_k^2 = 1 - \frac{k_{\max}}{k} \qquad (1.7)$$

where

k = estimated overdispersion parameter for the chosen model

k_{max}= the estimated overdispersion parameter for a model with only an intercept term.

Two other measures can be also used, the mean scaled deviance and the mean Pearson χ^2.

Taking up Bird and Hashim study again, before accident prediction models were created, it was necessary to define the individual variables to use in the analysis. Two main exposure variables were used, the length of the element, and the traffic flow on it. Other variables fell into two groups, direct geometric variables, and speed and consistency indices.

The first stage of the study involved the development of models for groups of accidents, for curves, tangents and then for the combined dataset. For all the models calibrated, side access density (direct variable) and the absolute difference between the speed limit and element operating speed (consistency variable), was found to be statistically significant.

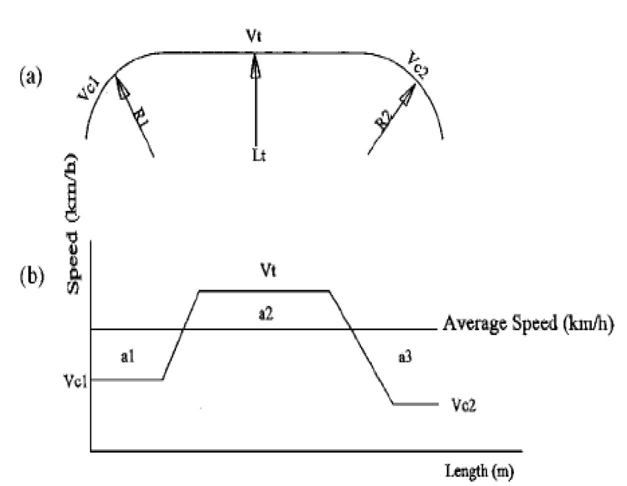
Review of Polus and Mattar-Habib (2004) [2.51]

These authors utilized operating speed profiles on nine two-lane rural roads segments having lengths ranging from 3 to 10 km each in northern Israel. Two measures of consistency were developed for these segments. The first measure was the normalized relative area (per unit length), bounded between the speed profile and the average speed line. The average operating speed, V_{avg}, was computed as the average weighted speed, by length, along the entire segment. If the areas bounded between the speed profile and the average operating speed line are denoted by a_i, as shown in Figure 1.3, then the first consistency measure is given as:

$$\text{Ra}=\left(\sum a_i\right)\Big/ L \tag{1.8}$$

Where *Ra* is the relative area (m/s) measure of consistency; $\sum a_i$ is the sum of areas bounded between the speed profile and the average operating speed (m^2/s); and *L* is the entire segment length.

The second measure of consistency was the standard deviation of speed (σ) along the road segment. The standard deviation is the most appropriate statistical measure of data distribution around the mean value. It was necessary to use this additional measure to complement the first measure because the *Ra* measure by itself provided similar result for somewhat different geometric characteristics in a few cases, though this was rare.



$$V_{avg} = \frac{\sum_{i=1}^{n} V_i * L_i}{\sum_{i=1}^{n} L_i}$$

$$Ra = \frac{\sum_{i=1}^{n} a_i}{L}$$

$$\sigma = \sqrt{\frac{(V_i - V_{avg})^2}{n}}$$

Figure 1.3. (a) Example of road section and (b) example of speed profile

The standard deviation of the operating speed was defined as:

$$\sigma = [(V_i - V_{avg})^2/n]^{0.5} \quad (1.9)$$

where σ is the standard deviation of the operating speeds (km/h); V_i is the operating speed along the i^{th} geometric element (tangent or curve) (km/h); V_{avg} is the average weighted (by length) operating speed along a road segment (km/h); and *n* is the number of geometric elements along a section (km/h).

These two measures of consistency provide an independent assessment of the resemblance (i.e., consistency) of speed performance along the entire road segment under study. Their main advantage is that they consider the consistency of the overall longitudinal segments, not just individual speed differentials between two successive elements.

As the relative weighted area bounded by the speed profile and the average weighted operating speed increases, so does the inconsistency of speeds. The standard deviation of operating speed also increases as the distribution becomes more dispersed. These two measures were found, opportunely combined in a road safety evaluation model developed by authors, significantly related to safety.

1.2.1. Key points

On the basis of the literature review presented so far, some key points need to be emphasized regarding statistical accident modelling for motorways, the focus of this thesis.

First, the generalized linear modelling technique (GLM) is appropriate for safety performance function calibration. This approach overcomes the limitations of conventional linear regression in accident frequency modeling, and allows a Poisson or Negative Binomial error structure, distributions that are more pertinent to accident frequency modelling, to be assumed. Together with this, new techniques of calibration can be used.

Second, the model form adopted has to account for the nonlinear relationship between accident frequency estimate and traffic volume variable; moreover it has to allow for the dependent variable to logically equal zero as the exposure variables goes to zero.

Third the independent variable definition and selection must account for road geometric features as well as consistency measures and road context-related characteristics.

Fourth, in parallel with the independent variable choice issue, the state of the art analysis revealed research focusing also on providing guidelines or suggestions to define and obtain homogeneous segments [1.10][1.14][1.21][1.38][1.51]. Cafiso et al. [1.52] also presented a detailed methodology purposely set up for dividing the entire path into segments characterized by homogeneous highway features related to safety.

On the basis of the literature review, and in order to organize the concepts relevant to this thesis, following is summarizes descriptions of the statistical regression technique used for safety performance function calibration, the choice of model form, and the most frequently used goodness of fit measures for safety performance function evaluation.

1.3. Regression Technique, Model Form and Goodness of Fit Evaluation

In the first instance a note on terminology is needed to clarify some terms used in the following chapters. The term accident prediction models, often used to indicate safety performance functions, usually denotes a multivariate model fitted to accident data in order to estimate the statistical relationship between the number of accidents and factors that are believed to be related to accident occurrence. The term “predictive” is somewhat misleading; “explanatory” would be a better term. Prediction refers to attempts to forecast events that are yet to occur, whereas accident prediction models are always fitted to historical data and can thus only describe, and perhaps explain, past events [1.53].

Moreover, the choice of the explanatory variables potentially affecting the safety performance of a site ought to be based on theory [1.54]. A theoretical basis for choosing explanatory variables might take the form of, for example, a causal model [1.55]. In practice, a theoretical basis for identifying explanatory variables is rarely stated explicitly [1.56]. The usual basis for choosing explanatory variables appears to be simply data availability. It is obvious that any analysis will be constrained by data availability. Nevertheless, the choice of explanatory variables should ideally not be based on data availability exclusively. Explanatory variables should include variables that:

- Have been focused in previous studies to exert a major influence on the number of accident;
- Can be measured in a valid and reliable way;
- Are not endogeneous, that is dependent on other explanatory variables included or on the dependent variable in the model.

Historically, two statistical modeling methods have been used to develop collision prediction models: conventional linear regression and generalized linear regression [1.57]. Recently however, generalized linear regression modeling (GLM) has been used almost exclusively for the development of collision prediction models. Several researchers (e.g. Jovanis and Chang 1986, Hauer et al. 1988, Miaou and Lum 1993) [1.58][1.9][1.59] have demonstrated the inappropriateness of conventional linear regression for modeling discrete, non-negative, and rare events such as traffic collisions. These researchers demonstrated that the standard conditions under which conventional linear regression is appropriate (Normal model errors, constant error variance, and the existence of a linear relationship between the response and explanatory variables) cannot be assumed to exist when modeling the occurrence of traffic collisions.

The GLM approach has the advantage of overcoming the limitations of conventional linear regression in accident frequency modeling. In particular, it allows also a Poisson or Negative Binomial error structure, distributions that are more pertinent to accident frequency modelling, to be assumed. The GLIM approach is described in the following [1.9][1.60][1.61].

Suppose Y_i is a random variable that describes the number of crashes at a given location *i* in a given period of time. Then Y_i is assumed to possess Poisson distribution and can be expressed as:

$$P(Y_i = y_i) = \frac{\lambda_i^{y_i} \cdot e^{-\lambda_i}}{y_i!} \quad (1.10)$$

where $P(Y_i=y_i)$ is the probability of occurring *y* crashes on the roadway section *i* in a given period of time and λ_i is the expected number of crashes on section *i* (i.e. $E(y_i)$). In addition, the mean or expected value of Y_i is assumed to be equal to its variance. That is:

$$E(Y_i) = Var(Y_i) = \lambda_i \quad (1.11)$$

where, $E(Y_i)$ is the expected number of crashes on section *i* and $Var(Y_i)$ is the variance of observed number of crashes. For a given set of explanatory variables (highway geometrics, traffic and other data), λ_i can be estimated using the formulation:

$$\ln(\lambda_i) = \beta X_i \tag{1.12}$$

where, X is a vector of explanatory variables and β is a vector of parameters to be estimated.

However, in some cases this method has limitations when applied to real world data due to the assumption of equal mean and variance. When the data are overdispersed or underdispersed (i.e. mean is note equal to the variance), use of this method would overestimate or underestimate the parameters. Many previous studies have found that crash data tend to be overdispersed in many situations with the variance being significantly higher than the mean. In such cases, any inferences made based on Poisson model estimations may lead to wrong conclusions. As a result of this, many researchers recommend using alternative methods in analyzing crash data, especially when the data is overdispersed. One such method is to utilize Negative Binomial (NB) distribution because it does not require the equal mean and variance assumption. In this method, the mean or expected value itself is assumed to be a random variable, which can be described by Negative Binomial distribution. In this case, λ_i can be now written as:

$$\ln(\lambda_i) = \beta X_i + \varepsilon_i \tag{1.13}$$

where ε_i is the unobservable error term with a gamma distribution. The variance of this distribution can be expressed as:

$$Var[Y_i] = \mu_i + k\mu_i^2 \qquad (1.14)$$

where μ_i is the expected number of crashes on section *i* and *k* is called the Negative Binomial dispersion parameter. The overdispersion occurs when the value of *k* is greater than 1 and when its value is zero, NB distribution reduces to Poisson distribution with which the variance is equal to the mean. The corresponding probability distribution under the NB assumption is given by:

$$P(Y_i = y_i) = \frac{\Gamma(y_i + 1/k)}{\Gamma(y_i + 1)\Gamma(1/k)} \cdot \frac{(k \cdot \mu_i)^k}{(1 + k \cdot \mu_i)^{(y_i + 1/k)}} \qquad (1.15)$$

where, $\Gamma(.)$ is the gamma function. The Negative Binomial model can be estimated using a maximum likelihood method to obtain the model parameters or β values and the dispersion parameter *k*. This can be carried out through maximization of the likelihood function (*L*), where, N is the total number of sections:

$$L(y_i, \mu_i, \alpha) = \sum_{i=1}^{N} \left[y_i \cdot \log(\alpha \cdot \mu_i) - (y_i + 1/\alpha)\log(1 + \alpha \cdot \mu_i) + \log\left(\frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \right) \right] \qquad (1.16)$$

The general form of the accident prediction model adopted is:

$$E(Y) = e^{a_0} \cdot L \cdot AADT^{a_1} \cdot e^{\sum_{j=1}^{m} b_j \cdot x_j} \qquad (1.17)$$

where:

E(Y)	=	expected accident frequency (accidents/time period);
L	=	length of the segment under consideration (km);
AADT	=	Average Annual Daily Traffic (AADT) (veh/day);
x_j	=	Any of m- additional variables;
a_0, a_1, and b_j	=	model coefficients;

This model form was selected because it is generally accepted as the form that better describes the accident phenomenon [1.10][1.21][1.38][1.51].

In particular, it logically estimates zero accidents if one of the two exposure variables (AADT or L) is equal to zero and it allows the non-linear relationship between traffic volume and accident frequency by means of a suitable calibration of the traffic volume coefficient.

Several measurements are usually used to assess the goodness-of-fit of the model and the significance of the model parameters [1.62]. These are the t-ratio for the model parameters, the scaled deviance (SD) and the Pearson χ^2 statistics for the model overall. The SD, defined as the likelihood ratio test statistic, estimated as twice the difference between the log likelihoods of the studied model and the full or saturated model where the full model has as many parameters as the observations, so that the model fits the data perfectly. Therefore, the full model, which possesses the maximum log likelihood achievable under the given data, provides a baseline for assessing the goodness-of-fit of an intermediate model with specified parameters. McCullagh and

Nelder [1.63] showed that for negative binomial error structure the scaled deviance is as follows:

$$SD = 2\sum_{i=1}^{n}\left[yi\ln(\frac{yi}{\hat{E}(yi)}) - (yi+k)\ln(\frac{yi+k}{\hat{E}(yi)+k}) \right] \tag{1.18}$$

where:

SD	=	scaled deviance;
y_i	=	observed number of accidents of the i^th segment;
$\hat{E}$ (y_i)	=	predicted number of accidents of the i^th segment;
k	=	negative binomial dispersion parameter;
n	=	number of the observations in the sample.

The Pearson χ^2 can be calculated by means of the following formula:

$$Pearson\ \chi^2 = \sum_{i=1}^{n}\frac{\left[y_i - \hat{E}(y_i)\right]^2}{Var(y_i)} \tag{1.19}$$

where:

Var(y_i) = variance of the i^{th} segment.

In order to verify the goodness-of-fit of the model, these two measures must be compared with the value obtained from the χ^2 distribution with the model's degrees of freedom and for a given level of significance. The model can be considered significant, if both measurements are less than the χ^2 critical value.

The Akaike's Information Criterion (AIC) was also used as goodness of fit measure [1.10][1.38][1.64]. The AIC value is calculated as follows:

$$AIC=-2\cdot logL+2\cdot p \qquad (1.20)$$

where:

logL = maximum log-likelihood of the fitted model;
p = number of parameters in the model.

The smaller the value of AIC, the better the model data fit. Therefore AIC can be used also to compare and rank different models calibrated on the same dataset.

Another measure of goodness of fit is R^2_k. This measure was calculated as follows [1.48]:

$$R_k^2 = 1 - \frac{k_0}{k} \qquad (1.21)$$

where:

k_0 = the overdispersion parameter estimated in the negative binomial model with only a constant term;
k = the negative Binomial overdispersion parameter estimated in the full model.

R^2_k is simple to calculate, and it yields a value between 0 and 1. Since it is based explicitly on the overdispersion parameter, it is especially applicable in evaluating negative binomial models. The higher the R^2_k value, the better is the fit.

An adjusted R^2_k is a modification of R^2_k that adjusts for the number of explanatory terms in a model:

$$R_k^2 adj = 1 - (1 - R_k^2) \cdot \frac{(n-1)}{(n-p-1)} \tag{1.22}$$

where

n = total number of observations in the sample;

p = total number of the parameters in the model (not counting the constant term).

A tool that can be adopted for the evaluation of the appropriateness of the model form chosen, is the cumulative residual analysis [1.56][1.65][1.66]. The residual for each data observation of the sample is equal to the difference between the observed and estimated values of the dependent variable. A standardized residual for the i^{th} observation, e.g., road segment, (SR_i) is computed from the following equation:

$$SR_i = \frac{(y_i - \hat{y}_i)}{\sqrt{\hat{y}_i + k \cdot \hat{y}_i^2}} \tag{1.23}$$

where:

y_i = observed number of accidents of the i^{th} segment;

$\hat{y}_i$ = estimated number of accidents of the i^{th} segment;

k = dispersion parameter

The cumulative standardized residual value for the j^{th} element is obtained as follow:

$$CSR_j = \sum_{i=1}^{j} SR_i \tag{1.24}$$

for j = 1, ,..., m and with m equal to the total number of the observation.

Cumulative standardized residuals analysis is carried out by plotting CSR_i versus exposure values, computed as the product of AADT and the length for each homogeneous section. The more the points of the CSR curve stay close to the x-axis, avoiding significant increase or decrease of the shape, the closer the estimates are to the observations and the more limited is the over/under-estimation phenomenon. Moreover, a CSR curve contained into the 2 standard deviation interval (e.g. 2σ), indicates the appropriateness of the model form chosen [1.67].

1.4. Empirical Bayes Estimation and the Role of the NB Dispersion Parameter

Ultimately, a safety performance function gives an estimate of the expected number of accidents for a roadway element that has a certain combination of traits. In most models, these include traffic volume and characteristics of highway geometry. Most safety performance functions will not include all factors that produce systematic variation in accident counts. Hence, estimates of the expected number of accidents derived from a safety performance function represent mean values for sites which have a given combination of traits. The expected number of accidents for a specific site will normally differ from the mean value for units which have similar general traits.

What is the best estimate of the long term expected number of accidents for a given roadway element, given the fact that some, but not all the factors affecting accident occurrence, are known? According to the empirical Bayesian method [1.68], the best estimate of safety is obtained by combining two sources of information:

1. The accident record for a given site;
2. A safety performance function showing how and how much various factors affect accident occurrence.

Let O be the observed number of accidents and E the normal, expected number of accidents estimated by a safety performance function. The best estimate (empirical Bayes estimate) EB of the expected number of accidents for a given site is given as follow:

$$EB = w \cdot E + (1 - w) \cdot O \qquad (1.25)$$

The parameter (empirical Bayes calculation weight) w determines the weight given to the estimated normal number of accidents for similar sites when combining it with the observed number of accidents in order to estimate the expected number of accidents for a particular site. Usually *w* is estimated as follow:

$$w = \frac{1}{1 + \frac{Var(E)}{E}} \qquad (1.26)$$

where Var(E) is the variance of the expected number of accidents E estimated by a safety performance function and is given as:

$$Var(E) = E + k \cdot E^2 \tag{1.27}$$

where k is the value of the dispersion parameter characterizing the safety performance function and estimated as the "shape-parameter" of the negative binomial distribution adopted for calibration in the regression model.

There is no doubt that the development of safety performance functions, or what we may term "modern" accident prediction models, during the past 15 years represents a major step forward in road safety research. Road safety research is now rapidly becoming a mature scientific discipline [1.69], a discipline that can be taught in universities and that provides basis for a rational approach to road safety management.

Development in the field of accident modelling has been so rapid, that some models that were considered as state of the art only ten years ago, look somewhat primitive today. But today there is a danger, as pointed out by Lord, Washington and Ivan (2005) [1.70], of moving too far in the direction of mathematical sophistication and perfect fitting of models. Accidents are very complex phenomenon; hence models also need to be complex in order to faithfully reproduce the main features of reality. Yet, the art of model building is, and will always be, the art of making the right simplifications. A good model is not necessarily an immensely complex model that perfectly fits the data in every detail. A good model is rather the simplest possible model that adequately fits the data, and that contains relationships that may be presumed to hold in general.

So as previously emphasized, safety performance functions (SPFs) are commonly calibrated using negative binomial regression in which a dispersion parameter that represents extra-Poisson variation is estimated. The negative binomial error structure is preferred to the conventional normal distributed structure assumed in conventional regression modeling because crash data are non-negative counts. The Poisson distribution is a special case of the negative binomial but is only applicable when all entities with identical values of the independent variables can be assumed to have identical means, a near impossibility because of the practical difficulty of accounting for all effects in crash prediction models. Indeed, the size of the calibrated dispersion parameter is a measure of how well a model represents the data and can be thus used to compare competing models for the same data.

Actually the primary use of the dispersion parameter is not in assessing goodness of fit, but in empirical Bayes (EB) estimation [1.71] in which the posterior mean of an entity with a known crash history is estimated as a weighted average of the crash counts and an SPF prediction. The dispersion parameter is used to calculate the relative weights for each component and is such that the smaller the parameter, the better the model is and the greater the weight assigned to it, relative to that assigned to the crash counts.

These EB estimates are used in safety management applications in such modern application tools as the Highway Safety Manual [1.71] and SafetyAnalyst [1.72], for which the two primary uses are:

1. To estimate the safety of sites for network screening to prioritize them for safety investigation;
2. The treatment evaluation to estimate the expected safety of a site had it not been treated.

It stands to reason, therefore, that the importance of precise estimation of the dispersion parameter should be established. This need is complicated by findings in a recent flurry of research papers that suggest that the dispersion parameter, contrary to earlier research, is not constant for a given data set but actually varies from site to site, depending on site characteristics such as segment length. Therefore the real question is: Does it matter that the dispersion parameter varies and, if so, does it matter how it varies? The question is especially topical since dispersion parameters in the Highway Safety Manual (HSM) and SafetyAnalyst either are constant or have a very simple form.

The varying form in the HSM and SafetyAnalyst is such that the dispersion parameter for certain classes of road segments is inversely proportional to segment length, as first suggested by Hauer [1.73] who argued logically that shorter segments have a higher accident frequency variance and consequently should have a higher dispersion parameter than longer segments, and that this variation should influence the long-term estimate of a segment's safety. Since this argument does not apply to intersection models, it is not surprising that dispersion parameters for these SPFs are constant in the HSM and SafetyAnalyst. However, others [1.74][1.75][1.76][1.77] have suggested that other variables can affect the dispersion parameter, and thus this parameter may also vary for entities such as intersections and rail-highway crossings. Sayed et al. [1.74], for example, in essence allowed the variation in the dispersion parameter for urban arterials to be accounted for by estimating the variance of the prediction as a function with a form similar to that used to estimate the mean. The dispersion parameter is then estimated as the square of the model prediction divided by the variance estimate. Although this approach does provide substantial flexibility in estimating the variation in the dispersion parameter, the model form for the variance is somewhat arbitrary and the logic of including terms other than length has not been established. Sayed et al. found that the varying dispersion parameter effectively increased the goodness of fit to the data by allowing for more modeling variability. However, in terms of model application to the identification and ranking of accident-prone locations, there were small or limited differences compared to using a constant dispersion parameter.

1.5. Chapter summary

The use of a reliable modeling approach is the first step to conduct reliable safety analysis.

In the present Chapter 1 a wide literature review on the Safety Performance Function is reported together with the state of the art methodology to address the regression to the mean effects. Meaningful review on modeling approach of different Authors are reported in the Chapter with their own comment on the related studies.

Various methodologies for the evaluation of the goodness of fit are described in the Chapter and their peculiarity and reliability.

References

[1.1]. Persaud B., Dzbik L., "Accident Prediction Models for Freeways", Transportation Research Record 1401, 1993.

[1.2]. Mahalel D., "A Note on Accident Risk", Transportation Research Record 1068, TRB, National Research Council, Washington D.C., 1985.

[1.3]. Hauer E., Lovell J., Persaud B., "New Directions for Learning About Safety Effectiveness", Report FHWA/RD-86-015, Federal Highway Administration and National Highway Traffic Safety Administration, January, 1986.

[1.4]. Mountain L., Fawaz B., Jarrett D., "Accident Prediction Models for Roads with Minor Junctions", Accident Analysis & Prevention, Vol. 28, No. 6, 1996.

[1.5]. UK DoT, "COBA Manual", Department of Transport, London: Assessment Policy and Methods Division, 1981.

[1.6]. Maycock G., Hall R. D., "Accidents at four-arm roundabout", Crowthorne, UK: Transportation Research Laboratory, LR 1120, 1984.

[1.7]. Pickering D., Hall R. D., Grimmer M., "Accidents at Rural T-Junctions", Crowthorne, UK: Transportation Research Laboratory, RR 65, 1986.

[1.8]. Hall R. D., "Accidents at Four-Arm Single Carriageway Urban Traffic Signals", Crowthorne, UK: Transportation Research Laboratory, CR 6, 1987.

[1.9]. Hauer E., Ng J., Lovell J., "Estimation of safety at signalized intersections", Transportation Research Record, 1185, Transportation Research Board, National Research Council, Washington D. C., 1988, pp. 48-61.

[1.10]. Abdel-Aty M. A., Essam Radwan A., "Modeling Traffic Accident Occurrence and Involvement", Accident Analysis & Prevention, Vol. 32, No. 5, 2000 , pp. 633-642.

[1.11]. Abdel-Aty M. A., Chen C., Brady E., "Analysis of Accident-Involvement Trends by Drivers Age in Florida", ITE Journal on The Web, February, 1999, pp. 69-74.

[1.12]. Mostofa H., "Modeling of Traffic Accidents on Principal Arterial", M.S. Thesis, Department of Civil Engineering, University of Central Florida, 1998.

[1.13]. Chen C., "Statistical Analysis of the Effect of Demographic and Roadway Factors on Traffic Crash Involvement", M.S.

Thesis, Department of Civil Engineering, University of Central Florida, 1997.

[1.14]. Garber N. J., Ehrhart A. A., “Effect of Speed, Flow, and Geometric Characteristics on Crash Frequency for Two-Lane Highways”, Transportation Research Record 1717, Paper No. 00-0458, 2000.

[1.15]. Pfundt K., “Three Difficulties in the Comparison of Accident Rates”, Accident Analysis and Prevention, Vol. 1, No. 3, 1969, pp. 253-259.

[1.16]. Gwynn D. W., “Relationship of Accident Rates and Accident Involvements with Hourly Volumes”, Traffic Quarterly, Vol. 21, No. 3, July 1967. pp. 407-418.

[1.17]. Hakkert A. S., Hocherman I., Mensah A., “Levels of Safety on Interurban Roads”, Transportation Research Record 1553, TRB, National Research Council, Washington, D. C., 1996, pp. 95-102.

[1.18]. Hall J. W., Pendleton O. J., “Rural Accident-Rate Variations with Traffic Volume”, Transportation Research Record 1281, TRB, National Research Council, Washington, D. C., 1990, pp. 62-70.

[1.19]. Brodsky H., Hakkert A. S., “Highway Accident Rates and Rural Travel Densities”, Accident Analysis and Prevention, Vol. 5, No. 1, 1983, pp. 73-84.

[1.20]. "Speed and Accidents", Vol. 2, Research Triangle Institute, Durham, N. C., 1970.

[1.21]. Pardillo Mayora J. M., Llamas Rubio R., "Relevant Variables for Crash Rate Prediction in Spain's Two Lane Rural Roads", Transportation Research Board 82nd Annual Meeting, Washington D.C., January, 2003.

[1.22]. Maher MJ., Summersgill I. A., "Comprehensive Methodology for the Fitting of Predictive Accident Models", Accident Analysis and Prevention 28 (3), pp. 281-96, Pergamon Press London, 1996.

[1.23]. Miaou S. P., "The relationship between truck accidents and geometric design of road sections: Poisson vs. negative binomial regressions", 73rd Annual Meeting TRB National Research Council, Washington D. C., 1994.

[1.24]. Vogt A. Y., Bared J., "Accident models for two-lane rural roads: segments and intersections", Report DOT-RD-75-196, FHWA, U.S. Department of Transportation, 1998.

[1.25]. Krammes R. et al., "Horizontal Alignment Design Consistency for Rural Two-Lane Highways", Report FHWA-RD-94-034, FHWA, U.S., Department of Transportation, 1995.

[1.26]. Lamm R. et al., "Highway Design and Traffic Safety Engineering Handbook", McGraw- Hill, New York, 1999.

[1.27]. Resende P., Benekohal R., “Effects of Roadway Section Length on Accident Modeling”, Traffic Congestion and Traffic Safety in the 21st Century Conference, ASCE, Chicago, IL, 1997.

[1.28]. Ng J. C. W., Sayed T., “Effect of Geometric Design Consistency on Road Safety”, NRC Research Press, Web Site http://cjce.nrc.ca, Canada, March, 2004.

[1.29]. Fitzpatrick K., Anderson I., Green P., Krammes R., Poggioli B., “Evaluation of design consistency methods for two-lane rural highways, executive summary”, Report FHWA-RD-99-173, Federal Highway Administration, U. S. Department of Transportation, Washington, D.C., 2000.

[1.30]. Poe C. M., Tarris J. P., Mason J. M. Jr., “ Relationship of operating speed to roadway geometric design speed”, Report FHWA-RD-96-024, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 1996.

[1.31]. Lamm R., Psarianos B., Mailaender T., “Highway design and traffic safety engineer handbook”, McGraw-Hill Companies, Inc., New York, N.Y., 1999.

[1.32]. Morral J., Talarico R. J., “Side friction demanded and margins of safety on horizontal curves”, Transportation Research Record 1435, Transportation Research Board, National Research Council, Washington D.C., 1994, pp. 145-152.

[1.33]. Polus A., Fitzpatrick K., Fambro D. B., "Prediction operating speeds on tangent section of two-lane rural highways", Transportation Research Record 1737, Transportation Research Board, National Research Council, Washington D.C., 2000, pp. 50-57.

[1.34]. Messer C. J., "Methodology for evaluating geometric design consistency", Transportation Research Record 757, Transportation Research Board, National Research Council, Washington D.C., 1980, pp. 7-14.

[1.35]. Krammes R. A., Glascock S. W., "Geometric inconsistencies and accident experience on two-lane rural highways", Transportation Research Record 1356, Transportation Research Board, National Research Council, Washington D.C., 1986, pp. 1-10.

[1.36]. Wooldridge M. D., Bauer K., Green P., Fitzpatrick K., "Comparison of driver visual demand in test track, simulator, on-road environments", Proceedings of the 79th Annual Meeting of the Transportation Research Board, National Research Council, Washington, D.C., CD-ROM, National Academy Press, Washington D.C., 2000.

[1.37]. Hassan Y., Sayed T., Tabernero V., "Establishing practical approach for design consistency evaluation", ASCE Journal of Transportation Engineering, 127(4), 2001, pp. 295-302.

[1.38]. Zhang C., Ivan J. N., Effects of Geometric Characteristics on Head-On Crash Incidence on Two-Lane Roads in Connecticut. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908*, TRB, National Research Council, Washington, D.C., 2005, pp. 159-164.

[1.39]. Wang Y., Nihan N., "Estimating the Risk of Collisions Between Bicycles and Motor Vehicle at Signalized Intersections", Accident Analysis and Prevention, Vol. 36, 2004, pp. 313-321.

[1.40]. Shankar V., Mannering F., Barfield W., "Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies", Accident Analysis and Prevention, Vol. 27, 2005, pp. 371-389.

[1.41]. Miaou S. P., "The Relationship between Truck Accidents and Geometric Design of Road Section: Poisson Versus Negative Binomial Regressions", Accident Analysis and Prevention, Vol.26, 1994, pp. 471-482.

[1.42]. "SAS online documents", SAS Institute Inc. Version 8 February 2013.

[1.43]. Ben-Akiva M., Lerman S. R., "Discrete Choice Analysis: Theory and Application to Travel Demand", The MIT Press, Massachusetts, 1985.

[1.44]. Bird R., Hashim I., "Exploring the Relationship between Safety and Consistency of Geometry and Speed on British

Roads", Transportation Research Board, 85th Annual Meeting, Washington D.C., January, 2006.

[1.45]. Anderson I. B., Bauer M., Harwood W., Fitzpatrick K., "Relationship to Safety of Geometric Design Consistency Measures for Rural Two-Lane Highways", Transportation Research Record: Transportation Research Board, No. 1658, TRB, National Research Council, Washington, D.C., 1999, pp.43- 51.

[1.46]. Hassan Y., Misaghi P., Awatta M., "Speed-based measures for evaluation of design consistency on Canadian roads", 3rd International Symposium on Highway Geometric Design, CD-ROM, Transportation Research Board, National Research Council, Washington, D.C., 2005.

[1.47]. Miaou S. P., Lu A., Lum H. S., "Pitfalls of Using R2 to Evaluate Goodness of Fit of Accident Prediction Models", Transportation Research Record: Journal of Transportation Research Board, No. 1542, National Research Council, Washington, D.C., 1995, pp.6- 13.

[1.48]. Miaou S. P., "Measuring the Goodness of Fit of Accident Prediction Models", FHWA-RD-96-040, Federal Highway Administration, Washington, D.C., 1996.

[1.49]. Fridstrom L., Ifver J., Ingebrigtsen S., Kulmala R., Thomsen L. K., "Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident

Counts", Accident Analysis and Prevention, Vol. 27, No. 1, 1995, pp. 1-20.

[1.50]. Agresti A., "An Introduction to Categorical Data Analysis", John Wiley, New York, 1996.

[1.51]. Cafiso, S., Di Silvestro, G. (2011). Performance of safety indicators in identification of black spots on two-lane rural roads, Transportation Research Record. Vol. 2237,pp. 78-87.

[1.52]. Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., Accident Prediction Models for the Evaluation of Safety Performance on Two Lane Rural Highways. 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.

[1.53]. Elvik, R., State-of-the-art approaches to road accident black spot management and safety analysis of road networks. TØI Report 883, Oslo, 2007.

[1.54]. Fridstrøm, L., J. Ifver, S. Ingebrigtsen, R. Kumala and L. K. Thomsen, Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accident Analysis and Prevention, 27, 1-20, 1995.

[1.55]. Asher, H. B., Causal modelling, Sage Publications, Beverly Hills, CA, 1976.

[1.56]. Hauer, E., Statistical road safety modelling, in Transportation Research Record: Journal of the

Transportation Research Board, No. 1897, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 81–87.

[1.57]. Sayed T., de Leur P., “Appendix 2F – Evaluation of Road Improvement Program”, Road Improvement Program, Evaluation Report, 2001, pp. 51-56.

[1.58]. Jovanis P. P., Chang H. L., “Modeling the relationship of accidents to miles traveled”, Transportation Research Record, 1068, Transportation Research Board, National Research Council, Washington D.C., 1986, pp. 42-51.

[1.59]. Miaou S., Lum H., “Modeling vehicle accident and highway geometric design relationships”, Accidents Analysis and Prevention, Vol. 25, No. 6, 1993, pp. 689-709.

[1.60]. Kulmala, R., “Safety at rural three- and four-arm junctions. Development of accident prediction models”, Espoo 1995, Technical Research Centre of Finland, VTT 233, 1995.

[1.61]. Dissanayake S., Ratnayake I., “Statistical Modelling of Crash Frequency on Rural Freeways and Two-Lane Highways using Negative Binomial Distribution”, Advances in Transportation Studies an inter-national Journal Section B 9, February, 2006.

[1.62]. Sutterhwaite S. A survey of research into relationship between traffic accidents and traffic volumes.

Supplementary report 692. Transport Research Laboratory. Crowthorne Berks, U.K. 1981.

[1.63]. McCullagh, P., and Nelder, J. A. *Generalized linear models*. Chapman and Hall, New York, 1989.

[1.64]. Montella, A., Colantuoni, L., and Lamberti, R. Crash Prediction Models for Rural Motorways. In Transportation Research Record: Journal of the Transportation Research Board, No. 2083, Washington, D.C., 2009, pp. 180-189.

[1.65]. Lord, D., and B. N. Persaud. Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure. In Transportation Research Record: Journal of the Transportation Research Board, No. 1717, TRB, National Research Council, Washington, D.C., 2000, pp. 102–108.

[1.66]. Cafiso, S., D'Agostino, C.. Safety Performance Function for Motorways using Generalized Estimation Equations. Procedia - Social and Behavioral Sciences Volume 53, 3 October 2012, Pages 901–910.

[1.67]. Hauer, E., Bamfo, J. Two tools for finding what function links the dependent variable to the explanatory variables. Published in Proceedings of ICTCT 97 Conference, November 5-7 1997, Lund, Sweden.

[1.68]. Hauer, E., Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic

Engineering Measures on Road Safety, Pergamon Press, Elsevier Science Ltd., Oxford, U.K., 1997.

[1.69]. Hauer. E., The road ahead. Journal of Transportation Engineering, 131, 333-339, 2005.

[1.70]. Lord, D., S. P. Washington and J. N. Ivan, Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention, 37, 35-46, 2005.

[1.71]. Highway Safety Manual (HSM). First Edition. FHWA American Association of State Highway and Transportation Officials. U.S.A. 2010.

[1.72]. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites, White Paper for Module 1—Network Screening, December 2002. http://www.safetyanalyst.org/whitepapers/module1.pdf

[1.73]. Hauer, E. Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. Accident Analysis & Prevention, Vol. 33, No. 6, 2001. pp. 799-808.

[1.74]. El-Basyouny, K., and T. Sayed. Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models. Transportation Research Record: Journal of the Transportation Research Board, No. 1950, Washington, D.C., 2006, pp. 9–16.

[1.75]. Miranda-Moreno, L. F., Fu, L., Saccomanno, F. F., and A. Labbe. Alternative Risk Models for Ranking Locations for Safety Improvement. Transportation Research Record: Journal of the Transportation Research Board, No. 1908, Washington, D.C., 2005, pp. 1–8.

[1.76]. Lord, D., and P.Y. Park. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. Accident Analysis & Prevention, Vol. 40, 2008, pp. 1441-1457.

[1.77]. Geedipally, S.R., D. Lord, and B.-J. Park. Analyzing Different Parameterizations of the Varying Dispersion Parameter as a Function of Segment Length. Transportation Research Record: Journal of the Transportation Research Board, No. 2061, Washington, D.C., 2008, pp. 46-54.

CHAPTER 2

HOW TO ADDRESS THE TIME TREND EFFECTS IN THE CALIBRATION OF SPFs

2.1. Introduction

Safety Performance Functions (SPFs) are useful tools for estimating the expected number of crashes over a road network which are typically used in the screening of sites with promise for safety improvements. In the present Chapter 2 a procedure of analysis for motorways network offering a comparison between the conventional analytical techniques based on GLM (Generalized Linear Model) largely described in Chapter 1 and a different approach based on General Estimating Equation (GEE) is reported. The GEE model, incorporating the time trend, is compared in terms of results and reliability in the estimation with conventional models (GLM) that do not take into account the temporal correlation of accident data [2.1]. The analysis reported later in the Chapter, as well as the great part of the references are based on a study of Cafiso and D'Agostino presented at 5th SIIV International Conference in 2012 in Rome and published on Procedia Elsevier - Social and Behavioral Sciences, [2.2]. As it will be clear later, the time trend effect in the reliability of SPFs, and in the evaluation of the effects of a treatment as well, plays a fundamental role above all when Motorways are analyzed where the crash rate is generally low and more years of analysis need to have an adequate reliability.

2.2. The time correlation effects in the SPF

Accident counts observed at a site (road segment, intersection, interchange) are commonly used as a fundamental indicator of safety performing road safety analysis methods. Because crashes are random events, crash frequencies naturally fluctuate over time at any given site. The randomness of accident occurrence indicates that short term crash frequencies alone are not a reliable estimator of long-term crash frequency. This year-to-year variability in crash frequencies adversely affects crash estimation based on crash data collected over short periods. The short-term average crash frequency may vary significantly from the long-term average crash frequency. This effect is magnified at study locations with low crash frequencies where changes due to variability in crash frequencies represent an even larger fluctuation relative to the expected average crash frequency. When a period with a comparatively high crash frequency is observed, it is statistically probable that the following period will be followed by a comparatively low crash frequency. This tendency is known as regression-to-the-mean (RTM), and also is evident when a low crash frequency period is followed by a higher crash frequency period.

Failure to account for the effects of RTM introduces the potential for “RTM bias”, also known as “selection bias”. Selection bias occurs when sites are selected for treatment based on short-term trends in observed crash frequency. RTM bias can also result in the overestimation of the effectiveness of a treatment (i.e., the change in expected average crash frequency). Without accounting for RTM bias, it is not possible to know if an observed reduction in crashes is due to the treatment or if it would have occurred without the modification. In light of what has been said the safety level of a site can not be simply defined by its accidents history. Generalizing it can be said that the safety of an element of a road must be defined by the average expected number of accidents in a long period of time. To this aim, the use of longer periods of observation would be more appropriate. In general, this period of analysis depends on the availability of both traffic and crash data, but in literature numerous studies have shown that periods longer than 5 years of investigation could reduce the accuracy in the estimation of the Safety Performance Function as they introduce the natural time trend that with the traditional analysis technique using generalized linear models (GLM) can not be taken into account. This phenomenon is very pronounced in the motorway sector, as accident rate is very low if compared to the urban or rural highways and a typical period of analysis that is enough on other contexts is not sufficient. If several years of analysis are available it is possible to take into account the annual variation or trend in the calibration of SPFs due to the influence of factors which change over time using the General Estimating Equation (GEE) that incorporates time trend [2.2]. In order to assess this variation, the number of accidents of each year is treated as a single observation. Unfortunately, this procedure generates the disaggregation of data by creating a temporal correlation that can not be identified with conventional procedures of model calibration using GLM [2.3][2.4][2.5].

Basing on the previous considerations, the objective of this study is to illustrate the application of the GEE procedure to traffic-safety studies when several years of data are available and when it is desirable to incorporate trend. The application is for a segment of a Motorway in Sicily, Italy, using data for the years 2003 through 2009. The GEE models with trend are compared with GLM that do not account for temporal correlation in the accident count data. It is necessary, first, to provide some background on accident modeling before introducing the GEE concept.

2.3. The Generalized Estimating Equation (GEE) methodology of calibration

Crash observed at a site i in the year t (Yi,t) are typical time series data across years and can, therefore, be represented by the following simplified model structure:

$$Y_{i,t} = \text{trend} + \text{regression term} + \text{random effects} + \text{local factors} \quad (2.1)$$

where "trend" refers to a long-term movement due to a change in the risk factors with time, the "regression term" is of the same form as the Safety Performance Functions (SPF), "random effects" accounts for latent variables across the sites, and the "local factors" refers to the dispersion between the normal safety level for similar locations and the safety level for the specific site. This last term indicates the effects of local risk factors and also general factors that are not included in the safety performance function. Random effects and local factors both contribute to the dispersion of crash counts as compared to the mean value estimated by the regression term.

The use of the Negative Binomial (NB) distribution to represent the distribution of crash counts is commonly accepted. Therefore, excluding trend effects (i.e. the phenomenon is stationary) GLMs are especially useful in the context of traffic safety, for which the distribution of accident counts in a population often follows the negative binomial distributions [2.6][2.7].

The important property of the GLM is the flexibility in specifying the probability distribution for the random component [2.8][2.9][2.10].

Many model forms exist for SPFs, but one of the most common ones is the following:

$$E\{\kappa\} = \alpha F^{\beta} \tag{2.1}$$

or the linear version:

$$\ln E\{\kappa\} = \ln \alpha + \beta \ln F \tag{2.2}$$

where:

- $E\{\kappa\}$: The expected number of accidents per unit of time;
- F: Traffic flows (e.g., vehicles/day, vehicles/hour); and
- α, β: Coefficients to be estimated.

These coefficients are estimated by the maximum-likelihood procedure using a variant of the Newton-Raphson method [2.11].

The conventional application of GLMs to estimate SPFs without trend is well developed and for traffic-safety applications, it is desirable to estimate different coefficients for each year for which data is available. For logical reasons, it is usually assumed that the β's are constant from year to year, and it is therefore only necessary to estimate the different α's for each year. In the estimation of these α's, each annual accident count is an observation, which creates difficulty because these counts are correlated.

To appreciate the difficulty caused by temporal correlation, consider a simple example in which the model defined by Equation 2.1 is to be developed for longitudinal data for which accidents and traffic flows are available for different time periods (t) at intersections identified from i = 1 to I. The model is given by the following equation:

$$E\{\kappa_t\} = \alpha F_t^{\ \beta} \qquad (2.3)$$

where:

- $E\{\kappa_t\}$: The expected number of accidents per time period t;
- F: Traffic flows for Year t (e.g., vehicles/day, vehicles/hour); and

- α, β :Coefficients to be estimated

The GLM estimate of coefficients for Equation 2.3 is the solution to the following estimating equation:

$$\sum_{i=1}^{I} D_i^{'} V_i^{-1} (Y_i - \mu_i) = 0 \tag{2.4}$$

where:

- $\mu_i = g^{-1}(X_i \beta)$;
- $\beta = \alpha, \beta$, coefficients of the model to be estimated
- $V_i = \sigma^2 [(1-\rho) \cdot I + \rho \cdot J]$ (covariance matrix); and (2.5)
- $D_i = \frac{\partial \mu_i}{\partial \beta} = \begin{bmatrix} \frac{\partial \mu_{1i}}{\partial \alpha} & \frac{\partial \mu_{1i}}{\partial \beta} \\ \vdots & \vdots \\ \frac{\partial \mu_{ni}}{\partial \alpha} & \frac{\partial \mu_{ni}}{\partial \beta} \end{bmatrix}$ (2.6)

In Equation 2.5, I is the ni × ni identity matrix, J is the ni × ni matrix all of whose elements are 1, and ρ is the correlation coefficient between any two measurements at the same link i. Note that the data in this example are assumed to be uniformly correlated, but the reader should be aware that other types of temporal correlation also exist. See Diggle et al. [2.12].

The variance of the GLM estimate of β then becomes:

$$Var(\beta) = \sigma^2 \left[\sum_{i=1}^{I} D_i^{'} V_i^{-1} D_i \right]^{-1} = \sigma^2 \left[\sum_{i=1}^{I} D_i^{'} \{(1-\rho) \cdot I + \rho \cdot J\}^{-1} D_i \right]^{-1} \tag{2.7}$$

In Equation 5, if the observation is positively correlated (ρ > 0), which often occurs when the repeated accident counts for the same link are used, the variance of α and β will be increased by a factor ρ.

Thus, the variance will be underestimated if this correlation is ignored. More importantly, ignoring the temporal correlation also could have an impact on the proper selection of coefficients because some coefficients may be wrongly accepted as significant because of the underestimated variance. For example, one might conclude that the year-to-year differences in α are significant when they are not.

The coefficients of the GLM incorporating trend in temporally correlated data still may be estimated using the traditional maximum likelihood methods. However, the likelihood function is very complicated to define and solve. For instance, additional assumptions are routinely needed to specify the likelihood function of non-Gaussian data. And, even if these assumptions are valid, the likelihood often involves numerous nuisance parameters that must be estimated in addition to the explanatory variables. To overcome this difficulty, an alternative method known as the GEE procedure was proposed by Liang and Zeger [2.13] and Zeger and Liang [2.14].

The GEE procedure is classified as a multinomial analogue of a quasi-likelihood function. The estimate of the coefficients can be found with the same previous Equation 2.7, but the temporal correlation in repeated observations is described by a ni × ni matrix R(λ), where λ represents the type of correlation with λ = [λ1 , . . . ,λ n − 1] and λi = cor (Yit , Yik) for t, k = 1, . . . , n−1, t ≠ k, and ni is the number of subjects.

Therefore, the new covariance matrix now becomes:

$$V_i = \sqrt{A_i} \cdot R_i(\lambda) \cdot \sqrt{A_i} \quad (2.8)$$

where Ai is an ni × ni matrix with diag [V(μi1), . . . V(μiTi)]. The covariance matrix is given by 2.10:

$$\text{cov}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\sigma}^2 \left[\sum_{i=1}^{I} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \right]^{-1} \quad (2.9)$$

One can find the solution by simultaneously solving Equations 2.9 and 2.10 with the iterative reweighted least-squares method [2.15]. This method is required because the estimates of both β and λ need to be found.

To solve the GEE correctly, every element of the correlation matrix Ri has to be known. However, in many instances, it is not possible to know the proper correlation type for the repeated measurements. To overcome this drawback, Liang and Zeger [2.13] proposed the use of a “working” matrix $\hat{V}$ of the correlation matrix Vi which is based on the correlation matrix $\hat{R}_i$. The estimate of the coefficients is found with the following equation:

$$\sum_{i=1}^{I} \boldsymbol{D}_i' \hat{\boldsymbol{V}}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = 0 \quad (2.10)$$

The covariance matrix of Equation 2.9 is given by:

$$\operatorname{cov}\left(\hat{\boldsymbol{\beta}}\right)=\boldsymbol{\sigma}^2\left[\sum_{i=1}^{I}\boldsymbol{D}_i^{'}\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{D}_i\right]^{-1}\left[\sum_{i=1}^{I}\boldsymbol{D}_i^{'}\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{V}_i\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{D}_i\right]\left[\sum_{i=1}^{I}\boldsymbol{D}_i^{'}\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{D}_i\right]^{-1} \tag{2.11}$$

The proposed methodology in Equations 2.10 and 2.11 possesses one very useful property in that $\hat{\beta}$ nearly always provides consistent estimates of β even if the matrix Vi has been estimated improperly. Thus, the confidence interval for β will be correct even when the covariance matrix is specified incorrectly. Therefore, it is not necessary to, a priori, examine the type of temporal correlation (e.g., independent, dependent). Techniques on how to analyze and interpret autocorrelation can be found in books on time-series analysis such as those by Box and Jenkins [2.15] and Diggle [2.16]. One important drawback, however, comes with this property. To assume that $\hat{\beta}$ is the proper estimate of β, it is required that the observation for each subject be known and available. If missing values exist, the estimate of the coefficients may be biased. The extent of the bias is influenced by the type of missing values (e.g. random or informative). Note that, in the case of $\hat{V}_i = V_i$, Equation 2.11 becomes the covariance matrix of Equation 2.9.

2.4. Comparison between models with and without trend: a case study

In the following a case study about the comparison of models with and without trend is reported. The analysis reported below, as well as the great part of the references are based on a study presented at 5th SIIV International Conference in 2012 in Rome and published on Procedia - Social and Behavioral Sciences [2.2]

2.4.1, Data description and Models calibration

The SPFs proposed in this work are basically of two types. The first are based on the application of a model which the only explanatory variables are the Annual Average Daily Traffic (AADT) and the segment length (L) (exposure variables), while the second class of models in which in addition to the exposure variable there are a series of other variable depending to the physical characteristic of the road segment (multivariable models). Both classes of models will be analyzed using both Generalized Linear Model (GLM) and Generalized Estimating Equation (GEE). With the use of the latter finally the data is analyzed by incorporating in the model time trend.

The general form of SPF is shown in equation 2.12:

$$E(Y) = e^{\alpha} \times L \times AADT^{a_1} \times e^{\sum_{i=1}^{m} b_i x_i} \qquad (2.12)$$

where:

- xi= generic additional variable,
- m= number of generic additional variable,
- α, a1, bi= m +2 regression parameters.

The estimate will be zero if only at least one of the exposure variables (L and AADT) assumes a zero value. With regard to the model depending only on the traffic shall be considered null the additional xi and it is taken into account only the length of segment and AADT. The application of the proposed safety performance function is performed on segments of the A18 CT-ME from the years 2003 to 2009 excluding 2004 because during year 2004 the Agency had adequate safety barriers in different parts of infrastructure changing the homogeneity of the segments within the year of analysis. The whole dataset consists of 652 segments of variable length and more than 70 m. Accident taken into account are fatal and injury, for a total of 451 accident in six years of analysis. Segmentation was carried out to maintain all the variables constant within each segment. The choice (of the threshold of 70 m) to have segment longer than 70 m comes out from a balance between the number of residual accidents, after the elimination of shorter segment and the problem related to an appropriate statistical inference.

In the case of multivariable models the SPF assumes the form:

$$E(Y) = e^{a_0+\alpha_t} \times L \times AADT^{a_1} \times e^{a_2 \cdot Curv + a_3 \cdot Gd + a_4 \cdot Ril + a_5 \cdot Viad + a_6 \cdot RSH + a_7 \cdot dq} \quad (2.13)$$

where:

- Curv, Gd, Ril, Viad, RSH dq= additional variable,
- αt= the additional factor which take into account time trend,
- a0, a1, a2, …, a7= 8 regression parameters.

To each of them are associated all the information necessary for the application of the multivariable models that are:

- The Roadside Hazard (RSH), which assumes 6 possible values (from 1 to 6, in increasing order of potential hazard), defined as follows: first, we consider only the conditions of the outer margin, assigning 1 to the trench, 2 embankment with appropriate barriers, 3 to the viaduct with adequate lateral barrier, 4 embankment with the side dam is not adequate, 5 to the viaduct with adequate lateral barrier, and then this value is the sum 1 in the case in which the median barrier is not to norm. This parameter is variable over time, due to maintenance and replacement of the barriers made by the Agency. In particular, it is changed from 2004 onwards in viaduct;
- The slope of the grade downhill (Gd), this variable is significant in that it influences the speed of heavy vehicles and therefore determines the average speed of flows. It assumes the value zero in the uphill and assumes the value of the slope of the segment when it is down;
- The lack of cross slope of the sections analyzed (dq), analytically defined as the difference between the cross slope required according to the new Italian standard taking into account the radius of curvature and the type of road and what detected in situ;

- The variables related to the type of section (RIL and VIAD), are variables of type exclusive. They report the status of the homogeneous section. Those that were statistically significant are embankment and viaduct is instead discarded the trench. Therefore the variables described above assume the value 1 in the sections where it is embankment or viaduct and 0 in other segments;
- The variable relating to the curvature of the homogeneous road element (Curv) defined as the actual curvature of each element.

2.4.2. Results and model validation

The application of the models leads to the results shown in Table 2.1, Table 2.2 and Table 2.3. These coefficients were estimated using the SAS software package [2.17] linearizing the equation and reporting for each estimated coefficient the standard error and for each model the relative dispersion parameter. The dispersion parameter represents the error in the construction of the model and is estimated by an iterative procedure using SAS software package [2.17]. The models are all over 6. The first two were calibrated using the GLM in particular the multivariate model is the model 1 and the second (model 2) is the basic model in which the only independent variable are AADT and L. The model 3 and 4 were calibrated with GEE taking into account the time trend. In particular, model 3 is the multivariate and model 4 depends only of

AADT and L. Finally, models 5 and 6 were calibrated using the GEE but not considering the time trend. Comparing the results obtained by evaluating the standard error and the value of the coefficients of the regression is shown as the model 1 and model 2 have identical coefficients respectively to the models 5 and 6. This is due to the fact that the dataset used to calibrate is identical for all models. What varies is the value of the standard error of coefficients which is generally greater in models calibrated with GEE. Underestimating the standard could have an impact on the proper selection of coefficients because some coefficients may be wrongly accepted as significant because of the underestimated variance.

Table 2.1. Estimates of the coefficients , (Standard Error) and [p-value] for the GLM Models (1-2)

		GLM	
		Multi (1)	**Mono (2)**
a_0	**Interc.**	-23.2595 (1.7521) [<.0001]	-20.1118 (1.5198) [<.0001]
a_1	**AADT**	1.6738 (0.1812) [<.0001]	1.3312 (0.1560) [<.0001]
a_2	**Curv**	0.6044 (0.1550) [<.0001]	--
a_3	**Gd**	15.2622 (4.6726) [<.0011]	--
a_4	**Ril**	0.5548 (0.2765) [.0448]	--
a_5	**Viad**	0.4192 (0.2201) [.0500]	--
a_6	**RSH**	-0.1846 (0.0916) [.0440]	--
a_7	**dq**	-16.0828 (5.1129) [.0017]	--
	Disp.	**0.7068 (0.2240)**	**0.8315 (0.2406)**

On the other hand if dispersion parameter is considered, it is noted that for the classes of models calibrated not considering the time trend both they are calibrated using the GLM than the GEE it is identical. This again depends on the fact that the database used is the same.

Table 2.2. Estimates of the coefficients , (Standard Error) and [p-value] for the GEE Models with Trend (3-4)

		GEE With Trend	
		Multi (3)	**Mono (4)**
a_0	**Interc.**	-23.3331 (1.7560) [<.0001]	-20.1725 (1.5270) [<.0001]
a_1	**AADT**	1.6733 (0.1815) [<.0001]	1.3285 (0.1563) [<.0001]
a_2	**Curv**	0.6051 (0.1555) [.0004]	--
a_3	**Gd**	15.1352 (4.6792) [.0018]	--
a_4	**Ril**	0.5600 (0.2761) [.0031]	--
a_5	**Viad**	0.4195 (0.2192) [.0050]	--
a_6	**RSH**	-0.1858 (0.0914)[.0036]	--
a_7	**dq**	-16.1112 (5.1164) [.0032]	--
α_t	**2003**	0.1247 (0.1785)	0.1062 (0.1799)
	2005	-0.0300 (0.1820)	-0.0205 (0.1832)
	2006	0.0788 (0.1772)	0.0914 (0.1784)
	2007	0.2495 (0.1706)	0.2676 (0.1719)
	2008	0.0354 (0.1788)	0.0472 (0.1801)
	2009	0.0000 (0.0000)	0.0000 (0.0000)
	Disp.	**0.7012 (0.2230)**	**0.8269 (0.2397)**

If the models with the time trend calibrated with GEE (models 3 and 4) are compared, with those that do not take into account time correlation it is evident that the time trend also intervene on the value of the coefficients of all variables in the model. This is explained by the fact that the database used for the calibration while being identical, it is interpreted by the GEE model as a repetition of the various segments for each year of analysis and that it varies in both the traffic that some of the considered parameters.

In general, the models which do not consider time trend analysis used the database as a whole do not taking into account the repetition of the segment in different years, therefore, considering a larger sample than model with time trend of t times where t is the number of years of analysis. This is the reason for the lower value of dispersion parameter in models that incorporate temporal trends than the other. In practical terms, analyzing the results reported in Table 2.1 through Table 2.3, it is possible to say that with the traditional models (models 1, 2, 5, 6) that incorporate temporal trends would underestimate the expected number of accidents in the years 2003, 2006 and 2007 while we overestimated the value of the 2005 expected number of accidents, all related to last year of analysis. Referring to the dispersion parameter, has to be noted that difference in the estimation of the dispersion parameter can produce different results when the Empirical Bayes procedure is applied [2.18].

The goodness of fit of the models generated was investigated using the method of cumulate residuals (CURE), i.e. the sum of the differences between the number of accident observed at a site and the expected value at the same site in the same year. As the residuals of all models are normally distributed with expected value equal to 0 and variance equal to σ [2.19], it is possible to calculate the variance of the expected value of such site as the square of the cumulate residuals. The purpose is to evaluate the variance of the system and the trend of the variation of AADT residuals in order to identify any abnormal deviations of the SPF used and evaluate how it fits to the dataset. The following figures show the cumulate residuals for the proposed models (Figure 2.1 through Figure 2.6).

Table 2.3. Estimates of the coefficients , (Standard Error) and [p-value] for the GEE Models without Trend (4-5)

		GEE Without Trend	
		Multi (5)	**Mono (6)**
a_0	**Interc.**	-23.2595 (1.7703) [<.0001]	-20.1118 (1.611) [<.0001]
a_1	**AADT**	1.6738 (0.1821) [<.0001]	1.3312 (0.1651) [<.0001]
a_2	**Curv**	0.6044 (0.1718) [.0004]	--
a_3	**Gd**	-15.2622 (4.8868) [.0018]	--
a_4	**Ril**	0.5548 (0.2599) [.0328]	--
a_5	**Viad**	0.4192 (0.2197) [.0500]	--
a_6	**RSH**	0.1846 (0.0887) [.0375]	--
a_7	**dq**	-16.0828 (5.4573) [.0032]	--
	Disp.	**0.7068 (0.2240)**	**0.8315 (0.2406)**

The advantage of the method CURE is that it is independent from the number of observations, which is the case with other methods on the goodness of fit of models [2.19][2.20]. Analyzing the graphs it can be concluded that, as expected, the multivariable models (1, 3, 5), even if it is plotted respect only AADT oscillates much more close to zero and tend to exceed the limits of 2σ only in the proximity of the ends. Further consideration should be made on the peaks of cumulate residuals for multivariable models. Indeed, in the multinomial model that consider temporal trends (3) both the positive and negative peaks of the cumulate residuals are generally smaller than in multi models that not incorporate temporal trends (1 and 5) showing a better fit of the model to the dataset. Finally also the lower amplitude of the oscillation of the residuals around the axis of abscissas indicates a lower dispersion of the residues to vary AADTs for multivariable models.

In practical terms, what just said for multivariable models translates into a best estimate of the expected number of accidents of the models that consider time trend with respect to the dataset used and therefore a better adaptation to reality. For the models in which the only explanatory variables is AADT (2, 4 and 6) there are no significant difference. It could be explained by the fact that AADT have only moderate variations in the analysis period.

In the present study were six different SPFs have been analyzed, two of which incorporate temporal trends in the calibration of models with the use of GEE. The remaining four SPFs were calibrated using basic and multivariable models by the way of classical GLM and GEE without trend.

The data used are relative to a section of motorway A18 for a total of 652 segments and considering a period of analysis from 2003 to 2009 with the exclusion of 2004. The purpose of the present study was to investigate the accuracy of different models that incorporate temporal trends respect to the classical models which do not take into account the temporal correlation of crash data. The time trend models were calibrated with the use of SAS software for which a script was created ad hoc by which it was possible to generate models that have a different constant value depending on the year of analysis. In this way it is possible capture the variations in the expected number of accidents of sites investigated, due to the time correlation, compared to a year of analysis included in the time period analyzed. By analyzing the results it can be stated that the time correction generates a better goodness of fit of models that incorporate temporal trends. It also corrects over-underestimation of the standard errors of the regression coefficients and of the dispersion parameter obtained by the more traditional GLIM approach. This results, even if derived by the example application on an Italian Motorway, are consistent with other studies [2.1] [2.21]. Improving accuracy in the calibration of regression parameters and standard errors improve the quality of the SPMs and lead to more refined results when the EB approach is used to control the phenomenon of regression to the mean. Another advantage is related to the possibility of using a broader period of analysis. In fact, GEE that incorporate temporal trends are not affected by the extension of the period of analysis for two reasons. The first is that the data are analyzed as repetitions of the variables in different years and therefore although this reduces the size of the sample, makes sure that the analyst can use more years of analysis. The second is that the temporal correlation that is generated between the sites in different years does not generate errors related to the type and size of the correlation matrix used in the calibration of the model and this allows to take into account longer analysis period. This characteristic is useful especially when long periods of observations are needed to increase the sample size or to carry out before/after studies. In contrast, the calibration of the models with GEE is critical in the case of missing values and therefore requires quality and detail of data greater than the traditional techniques [2.1][2.2].

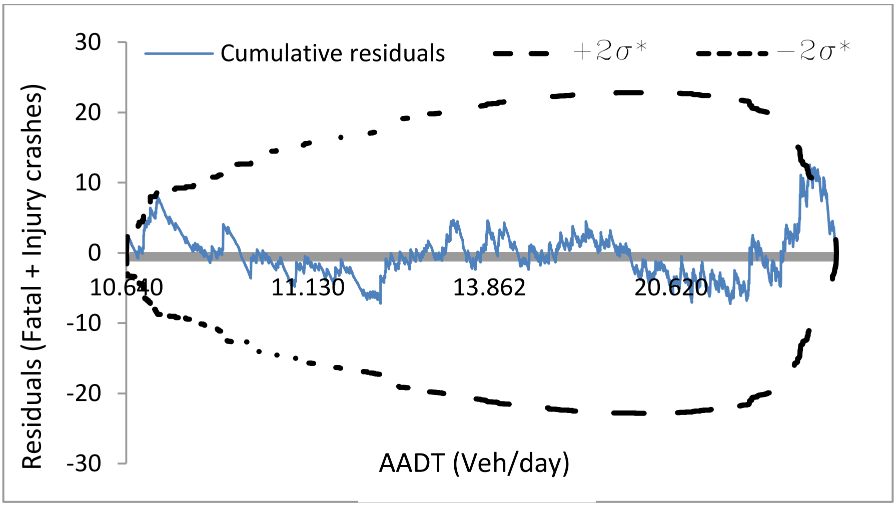


Figure 2.1. CURE Plots with ± 2σ for GLM model Multivariable (1)

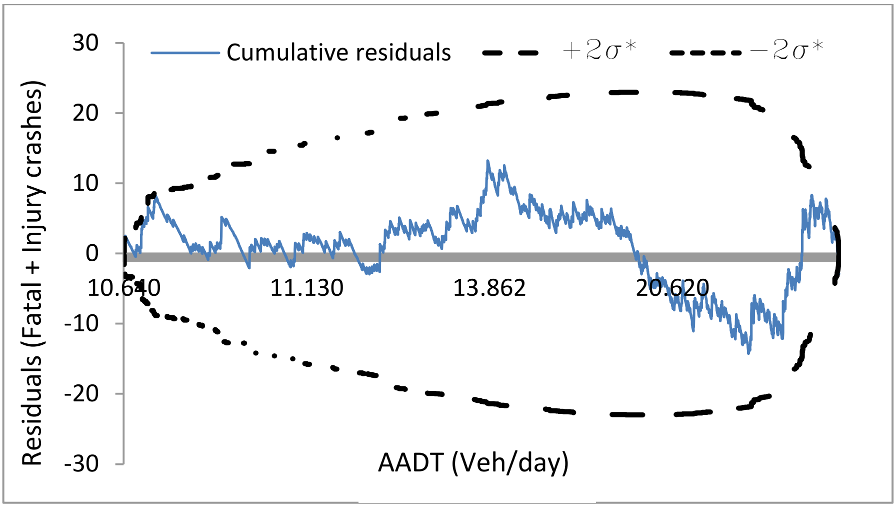


Figure 2.2. CURE Plots with ± 2σ for GLM model Monovariable (2)

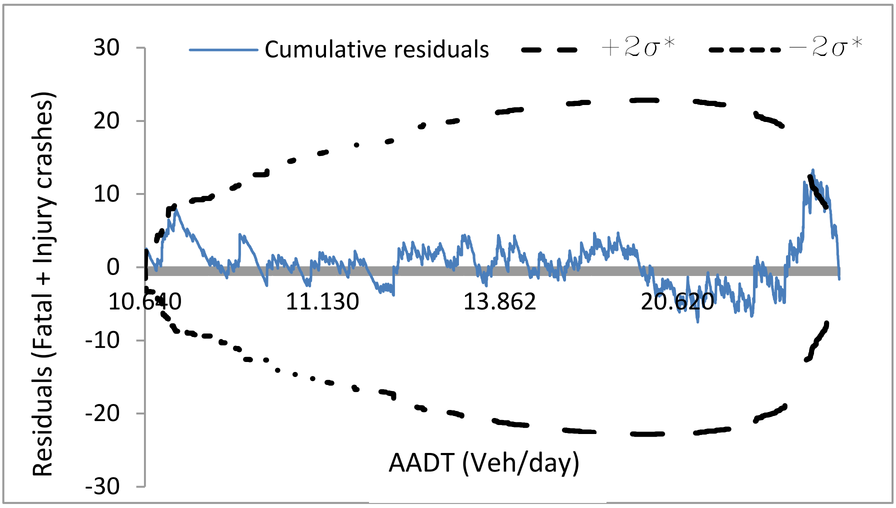


Figure 2.3. CURE Plots with ± 2σ for GEE model with trend Multivariable (3)

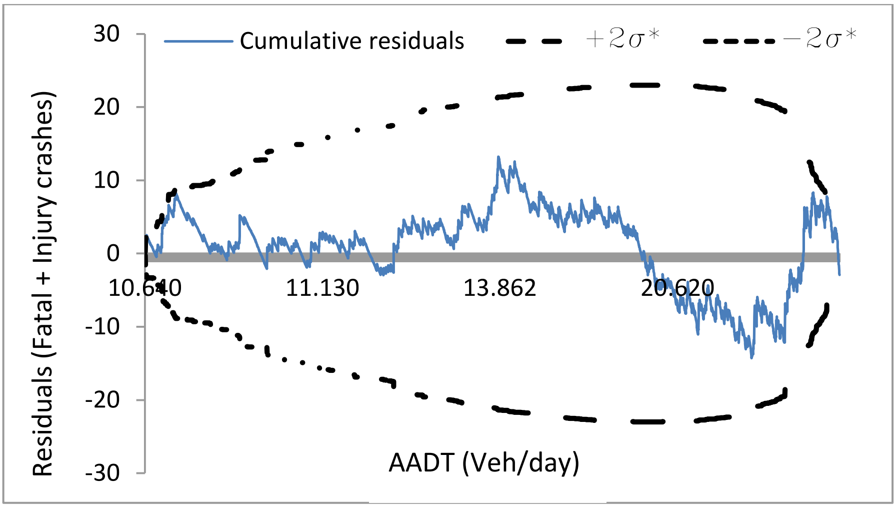


Figure 2.4. CURE Plots with ± 2σ for GEE model with trend Monovariable (4)

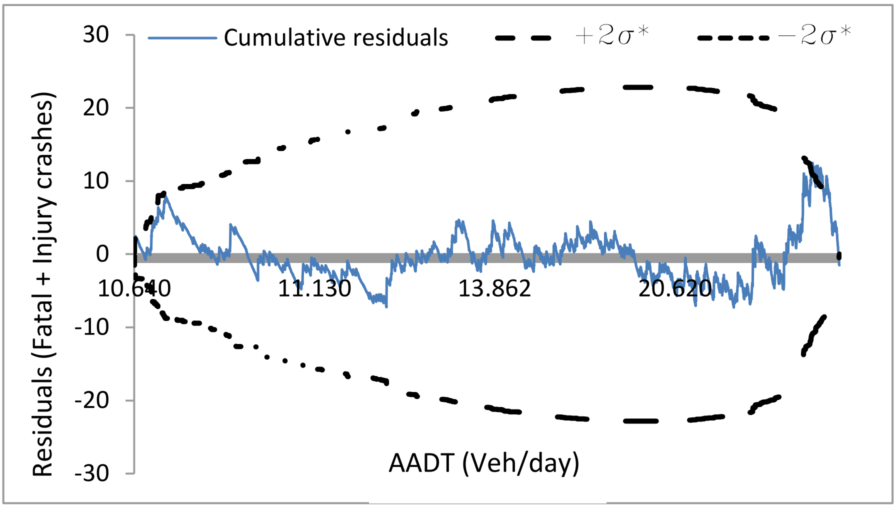


Figure 2.5. CURE Plots with ± 2σ for GEE model without trend Multivariable (3)

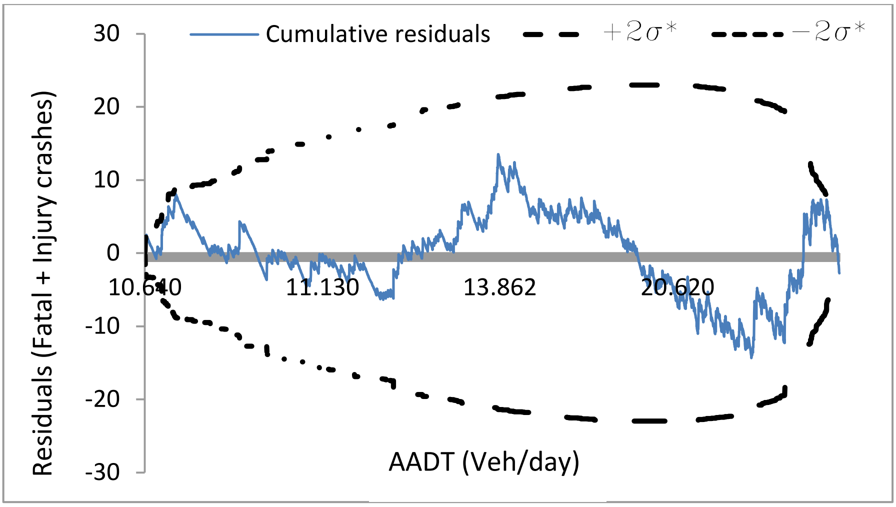


Figure 2.6. CURE Plots with ± 2σ for GEE model without trend Monovariable (6)

2.5. Chapter summary

Time trend effects play a fundamental role above all when motorways are analyzed. The low crash rate on comparison with others rural infrastructures means that to have reliable model longer period of analysis have to be taken into account.

A study presented at 5th SIIV International Conference in 2012 in Rome and published on Procedia Elsevier - Social and Behavioral Sciences [2.2] is reported on the topic. In that study a comparison between the traditional GLM approach and the GEE methodology of calibration was performed, analyzing the goodness of fit of the estimated models.

A higher reliability of the models which incorporate time trend and a general higher goodness of fit of the models which incorporates more variables is the conclusion of the Study.

References

[2.1]. Lord, D. and Persaud, B. N. (1998). Accident Prediction Models With and Without Trend Application of the Generalized Estimating Equations Procedure. In Transportation Research Record, No. 1717. 1.66

[2.2]. Cafiso, S., D'Agostino, C.. Safety Performance Function for Motorways using Generalized Estimation Equations Original Research Article Procedia - Social and Behavioral Sciences, Volume 53, 3 October 2012, Pages 900-909.

[2.3]. AASHTO (2011). Highway Safety Manual, pp. 105-106, Washington, D.C., USA. 1.71

[2.4]. Cafiso, S., Di Silvestro, G., Persaud, B., Begum, M.A. (2010). Revisiting variability of dispersion parameter of safety performance for two-lane rural roads.,Transportation Research Record, Vol. 2148, pp,38-46. 1.79

[2.5]. Cafiso, S., Di Silvestro, G. (2011). Performance of safety indicators in identification of black spots on two-lane rural roads, Transportation Research Record. Vol. 2237,pp. 78-87. 1.52

[2.6]. Kulmala, R. (1995). Safety at Rural Three- and Four-Arm Junctions: Development and Applications of Accident Prediction Models. VTT Publications 233, Technical Research Centre of Finland.1.61

[2.7]. Nicholson, A. and Turner, S. (1996). Estimating Accidents in a Road Network. In Proceedings of Roads 96 Conference, Part 5, New Zealand, , pp. 53–66. 1.80

[2.8]. Dunlop, D. D. (1994). “Regression for Longitudinal Data: A Bridge from Least- Squares Regressions”. The American Statistician, Vol. 48, No. 4, , pp. 299–303. 1.81

[2.9]. McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models: Second Edition. Chapman and Hall, Ltd., London. 1.64

[2.10]. Myers, R. (1990). Classical and Modern Regression with Applications. Duxbury Press, Belmont, U.K.. 1.82

[2.11]. Green, P. (1984). Iterative Reweighted Least Squares for Maximum-Likelihood Estimation and Some Robust and

Resistant Alternative (with discussion). Journal of the Royal Statistical Society, Series B, Vol. 46, pp. 149–162. 1.83

[2.12]. Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994). “Analysis of Longitudinal Data”. Clarendon Press, Oxford, U.K.. 1.84

[2.13]. Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. Biometrika, Vol. 73, , pp. 13-22. 1.85

[2.14]. Zeger, S. L. and Liang, K.Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. Biometrics, Vol. 42, pp.121-30. 1.86

[2.15]. Box, G. P. and Jenkins, G. M. (1970). Time-Series Analysis: Forecasting and Control (rev. ed.). Holden-Day, San Francisco, California. 1.87

[2.16]. Diggle, P. J. (1990). Time Series: A Biostatistical Introduction. Oxford University Press, Oxford, U.K.. 1.88

[2.17]. “SAS online documents”, SAS Institute Inc. Version 8 February 2013.

[2.18]. Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. Accident Analysis & Prevention, Vol. 33, No. 6,. pp. 799-808. 1.74

[2.19]. Hauer, E. and Bamfo, J. (1997). Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables. In Proceedings of ICTCT Conference, Lund, Sweden. 1.67

[2.20]. Miaou, S.-P. (1996). Measuring the Goodness-of-Fit of Accident-Prediction Models. Report FHWA-RD-96-040. FHWA, McLean, Va.. 1.48

[2.21]. Giuffrè O., Granà A., Giuffrè T., Marino R. (2007), Improving Reliability of Road Safety Estimates based on High Correlated Accidents Counts. Transportation Research Record, No. 2019, 2007.

CHAPTER 3

OPTIMIZATION OF THE GOF OF THE MODELS, VARYING THE SEGMENTATION APPROACH, FOR ROAD SAFETY SITE RANKING

3.1. Introduction

With the new European regulation on Road Safety, (96/2008 CE) and the Italian 35/2011, the transposition of the European code, the ranking of the hazardous location is central in the activity.

To have a reliable ranking of segment with high concentration of crashes, need to have a reliable SPF from a goodness of fit point of view, and a good methodology of calibration, able to correct the regression to the mean and time trend effects.

As shown in the previous Chapters Safety Performance Functions (SPFs) are crucial to science-based road safety management. Success in developing and applying SPFs depends fundamentally on two key factors: the validity of the statistical inferences for the available data and on how well the data can be organized into distinct homogenous entities. The latter aspect plays a key role in the identification and treatment of road sections or corridors with problems related to safety. Indeed, the segmentation of a road network could be especially critical in the development of SPFs that could be used in safety management for roadway types, such as motorways (freeways in North America), that have a large number of variables that could result in very short segments if these are desired to be homogeneous.

This consequence, from an analytical point of view, can be a problem when the location of crashes is not precise and when there is an over abundance of segments with zero crashes.

Lengthening the segments for developing and applying SPFs can mitigate this problem, but at a sacrifice of homogeneity. As it will be clear later the best results were obtained for the segmentation based on two curves and two tangents within a segment and the segmentation with fixed length. The segmentation characterized by a constant value of all original variables inside each segment was the poorest approach by all measures.

The great part of the elaboration related to the segmentation approach in the present Chapter 3 are based on a paper by Cafiso et al. [3.1] published on 92nd TRB annual meeting as well as the great part of the references which contributing in the Chapter.

At the end of the Chapter a ranking analysis is performed and compared with the Potential for Safety Improvement Index.

3.2. Methodological approach and literature review

Safety performance functions (SPFs) are crucial to science-based road safety management using, e.g., the methods prescribed in the Highway Safety Manual [3.2]. These functions are statistical models used to estimate the expected crash frequency for a facility [3.3] based on its characteristics, mainly traffic volume, which accounts for the majority of the variability in crash frequency, and geometric variables. These functions are developed from data for a number of similar sites. Success in development or application of an SPF for road segments depends strongly on how well the data can be organized into distinct homogenous entities, i.e., on the approach to segmentation.

Segmentation, when based on multiple variables, may lead to very short homogeneous segments [3.4]. For example, when using the segmentation approach proposed by the HSM, the presence of very short segments does not allow proper statistical inference for several reasons. The most important are the non-perfect identification of the location of crashes, which is often taken from police reports, [3.5], and the fact that crashes are rare events resulting in a great number of segments with zero crashes. Lengthening segments to avoid these issues will sacrifice homogeneity.

In the literature there are a number of different approaches to segmentation. Miaou and Lum suggested that short sections, less than or equal to 80 m could create bias in the estimation of linear models, but not when using Poisson models [3.6]. Similarly, Ogle et al. demonstrated that short segment lengths, less than 160 m, cause uncertain results in crash analysis [3.7]. Cafiso et al. [3.8] showed that to increase performance in identifying correct positives as black spots, segment length should be related to AADT with lower AADT values requiring longer segment lengths. Qin et al. studied the relationship between segmentation and safety screening analysis [3.5] using different lengths of sliding windows to identify hazardous sites, and concluded that short segments as well as those that are too long create a bias in the identification of sites with safety problem.

Some studies focused on the relationship between crashes and road geometry in addressing segmentation. For example, Cenek et al. [3.9], who investigated this relationship, for rural roads data, used a fixed segment length of 200 m. A similar study was made by Cafiso et al. [3.10] using homogeneous section with different lengths on a sample of Italian two lane rural roads, aggregating variable related to curvature and roadside hazard and concluding that model that contains geometry and design consistency variables are more reliable than others. Other studies suggested different ways to aggregate segment data to avoid lengths that are too short. For example, Koorey proposed the aggregation of curves and tangents when the radius of curves exceeds a predetermined threshold value [3.11].

The Highway Safety Manual (HSM) [3.1] recommends the use of homogeneous freeway segments with respect to AADT, number of lanes, curvature, presence of ramp at the interchange, lane width, outside and inside shoulder widths, median width and clear zone width. There is no prescribed minimum segment length for application of the predictive models for freeways, but there is a suggestion of a segment length not less than 0.10 miles.

Given the variety of approaches and the fact that there is no apparent preferred one, this study seeks to investigate alternative methodologies for segmentation, including the HSM procedures, using sample data from Italian motorways. All but one of these methods aggregate and redefine variables over longer segments while seeking to retain the geometric and exposure characteristics of the segment as best as possible. SPFs calibrated for different segmentations are compared in terms of goodness of fit and the variables captured. Stepwise regression was used for each of five different segmentation concepts to select the best combination of variables. In addition, for each segmentation concept, two simpler models were estimated and compared, a base model and curvature-based model that is described later.

3.3. Data gathering and treatment

The data used for this investigation are based on an Italian rural motorway, the "A18" Messina-Catania, which is approximately 76 km (47.2 miles) long. The cross section is made up from 4 lanes, 2 in each direction, divided by a median with barriers. The analysis period is for the 8 yeas from 2002 until 2009, during which 887 severe (fatal plus injury) crashes according to the official statistics on motor vehicle collisions provided by the Italian National Institute of Statistics (ISTAT) [3.12]. Table 3.1 shows basic statistics for the dataset used for analysis. In this study, only the road segments were analyzed; interchange data and the part of segment directly influenced by the presence of intersection were discarded. Every segment contiguous to an intersection starts from a distance of 50 m (164 ft) from the bevel for the insertion of the service lanes for exit from, and entry into the main flow. The data available, in addition to AADT (Figure 3.1), were: radius of curvature, vertical gradient, type of section, and roadside features (presence and typology of the lateral and median barriers). It was necessary to adapt segment data according to the various segmentations to make the variables significant.

Table 3.1. Details of the database used to estimate models.

Year	Range AADT	Crash (Fatal + Injury)	Length (km)
2002	8696 – 24904	94	145.08 (two directions)
2003	9082 -26123	95	
2004	9423 – 26947	100	
2005	10944 – 26882	104	
2006	7792 – 26414	113	
2007	7917 – 27001	119	
2008	7651 – 26783	113	
2009	9066 – 26743	93	
Total Crash		831	

In order to divide the sample into homogeneous segments, it was necessary to combine all the variables into a usable form, paying attention to the final form of the equation used for developing the SPFs for each segmentation approach [3.12].

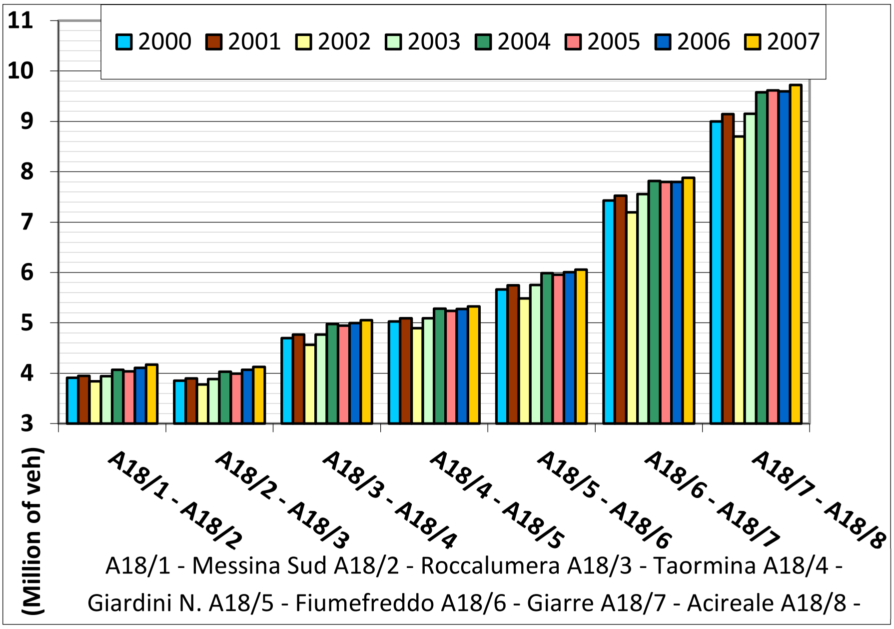


Figure 3.1. Details of annual traffic (millions of vehicles) during the analysis period.

The Average Annual Daily Traffic (AADT), which describes the exposure to crash risk, was not constant across some segments for two segmentation concepts; a length-weighted AADT was used as an approximation when this situation occurred. This approximation is more appropriate where a linear relationship between the dependent variable and AADT exists. (For the SPFs developed in this study, the crash-AADT relationship was approximately linear.) Other variables were similarly aggregated over segments in which they may not be constant for one or more segmentation approaches. The variables taken into account, apart from AADT, are related to geometry. The original data were: curvature, gradient of each segment and barrier condition. The aggregation of these data is described below:

- Curvature treated as curvature change rate (CCR) [3.13] of the segment, calculated as follows:

$$CCR = \frac{\sum_i |\gamma_i|}{L} [gon/m] \tag{3.1}$$

where γ_i is the deflection angle for a contiguous element (curve or tangent) i within a section of length L;

- Slope Change Rate (SCR) for the vertical profile of the road segment, which represents the variation of the slope inside a single segment, calculated as follows:

$$SCR = \frac{\sum_i |\delta_i|}{L} [gon/m] \tag{3.2}$$

where δ_i is the deflection angle for a slope related to the horizontal alignment within a section of length L;

- I, defined as the weighted average of the vertical gradient (up and down) with the reference length within each segment;
- Id, defined as the weighted average of the vertical gradient (down) with the reference length within each segment.
- Roadside Hazard (RSH) along a motorway, which is based both on type of section (trench, embankment, viaduct) and on the type of barrier with reference to the European standard EN 1317-1 1998) [3.14] [3.15] [3.16] [3.17]. RSH assumes 6 possible values (from 1 to 6, in increasing order of potential hazard), defined as follows: first, we considered only the conditions of the outer margin, assigning 1 to the

trench, 2 to embankment with appropriate barriers (complying with EN 1317-1), 3 to the viaduct with adequate lateral barrier, 4 to embankment where the side dam is not adequate, 5 to the viaduct with inadequate lateral barrier. A value of 1 was added if the median barrier is not adequate (Table 3.2). For this variable, the percentage of the length of a segment in which the RSH value was 6 (RSH6) or 5 and 6 (RSH56) was used;

- TUN, which indicates the percentage of the length of segment that is a tunnel;

Table 3.2. Value of RSH by type of section and condition of lateral and median barriers *(LA and MA indicate adequate lateral and median barriers; LN and MN indicate inadequate lateral and median barriers.*

	RSH for Lateral Barrier only		RSH for Median Barrier only		RSH for Lateral and Median Barrier combinations			
	LA	LN	MA	MN	LA MA	LA MN	LN MA	LN MN
Trench	1	1			1	2	1	2
Enban.	2	4	0	1	2	3	4	5
Viad.	3	5			3	4	5	6
Tunnel	4				4			

3.4. Analysis and results

In order to assess the influence of the organization of the data into segments on the goodness of fit of an SPF, five different segmentation approaches are assessed in this research, taking as reference the traffic (AADT) and the curvature. Specifically, these are:

- Segmentation 1: Homogeneous segments with respect to AADT and curvature, as suggested by HSM, using AADT and curvature as explanatory variables;
- Segmentation 2: Data organized to have within each segment 2 curves and 2 tangents, avoiding having short segments when using a single curve;
- Segmentation 3: Segments have constant AADT; other variables may not be constant.
- Segmentation 4: Segments have a constant length. Specifically, a length of 650 m was chosen, coinciding with the maximum length of an interchange area, and selected to be just longer than the longest horizontal curve. This length was chosen to minimize the problem of incorrect location of crashes on Italian motorways.
- Segmentation 5: All the variables used in the stepwise procedure are constant within each segment with their original value.

For the segmentation based on curvature and AADT, very short segments were eliminated in order to have segments with length more than 100 m. Using different segmentation approaches also changes the range of variation of the variables used to estimate the model. Table 3.3 and Table 3.4 show the range of the variables used for each segmentation approach. For Segmentation

5, characterized by the value of the original data constant inside each segment, it is not possible to use an aggregated variable for RSH and TUN, so these are used as categorical variables with their original value.

Table 3.3. Range (min-max) of variables for segmentation approaches 1,2 and 3.

	Seg_1 (Curve Based)	Seg_2 (2 Curves, 2 Tangents)	Seg_3 (AADT Based)
Length (m)	100.1 – 1563.4	234.7 – 3307.6	4882.3 – 21856.3
SCR (gon/m)	0 – 0.31	0 – 0.10	0.036 – 0.086
CCR (gon/m)	0 – 0.031	0 – 0.014	0.034 – 0.068
RSH6 (%)	0 – 70.23	0 – 55	0 – 10.61
RSH56 (%)	0 – 100	0 – 100	0 – 12.03
RSH	-	-	-
TUN (%)	0 – 100	0 – 75.4	0 – 49.1
I (Gon)	-0.042 – 0.045	-0.031 – 0.031	-0.0086 – 0.0088
Id (Gon)	0 – 0.043	0 – 0.031	0.0023 – 0.014

Table 3.4. Range (min-max) of variables for segmentation approaches 4 and 5.

	Seg_4 (Fixed Length)	Seg_5 (Homogeneous)
Length (m)	650.0	12 – 979.1
SCR (gon/m)	0 – 0.28	0 – 0.35
CCR (gon/m)	0 – 0.024	0 – 0. 33
RSH6 (%)	0 – 66.03	-
RSH56 (%)	0 – 100	-
RSH	-	1 - 6
TUN (%)	0 – 100	0 - 1
I (Gon)	-0.038 – 0.038	-0.042 – 0.045
Id (Gon)	0 – 0.038	0 – 0.043

The Generalized Estimating Equation (GEE) [3.18] method was used to estimate model coefficients, using the Statistical Analysis System (SAS) software package. The GEE procedure is classified as a multinomial analogue of a quasi-likelihood function, which allows the consideration of time trend in the models. Consistent with the state of research in developing these models, the negative binomial error distribution was assumed for the count of observed crashes [3.3].

The analysis of the road network using the predictive models includes the choice of a period of analysis. In general, this period of analysis depends on the availability of both traffic and crash data, but, in the literature, numerous studies have shown that periods longer than 5 years could introduce bias into the mathematical model for the variables linked to any physical changes of the network, or to the natural time trend, which, without the use of GEE, cannot be taken into account. The GEE procedure incorporates time trend, so is well suited to modeling data for long time periods. Specifically, it accounts for the temporal correlation that results when data for long periods are disaggregated into separate observations for each year.

To evaluate the goodness of fit (g.o.f.) of the models, two different methodologies were applied: the Quasilikelihood under the Independence model Criterion (QIC) [3.19] [3.20] and the Cumulative Residuals (CURE) plot [3.21]. The QIC statistic is analogous to Akaike’s Information Criterion (AIC) statistic used for comparing models fit with likelihood-based methods. Since the generalized estimating equations (GEE) method is not a likelihood-based method, the AIC statistic is not applicable. The QIC has the following form:

$$QIC = -2Q(\hat{\mu}; I) \tag{3.3}$$

where I represents the independent covariance structure used to calculate the quasi-likelihood and $\hat{\mu} = g^{-1}(x\hat{\beta})$ where $g^{-1}()$ is the inverse of link function.

When using QIC to compare two structures or two models, the model with the smaller statistic is preferred [3.19] [3.20] [3.22].The smaller the value of QIC, the better is fit of the model to the data. Therefore QIC can be used to compare and rank different models.

For the present study, another advantage of QIC is that the g.o.f. of models with different numbers of parameters can be compared.

The CURE method to evaluate the goodness of fit is based on the study of residuals, i.e., the difference between the number of crashes observed at a site and the expected value at the same site and in the same year. Assuming that residuals are normally distributed with expected value equal to 0 and a variance equal to σ (20), it is possible to calculate the variance of the expected value as the square of the cumulate residuals. The trend in the residuals with respect to AADT (or other variables) can be evaluated relative to the variance to qualitatively assess goodness of fit. The CURE method, is used to the examination of residuals after the estimation of the SPFs. Usually it can be used to examine whether the chosen functional form indeed fits the explanatory variable along the entire range of its values represented in the data. The plot of cumulative residuals should oscillate around 0, end close to 0, and not exceed the ±2σ* bounds. An upward/downward drift is a sign that the model consistently predicts fewer/more accidents than were counted. When the plot of residuals does not show any systematic drift, by examining the cumulative residuals, it can be assumed a good fitting of SPF to data.

The selection of the explanatory variables to be included in the model was made using a stepwise methodology inserting at first all the variables available, and testing for each of the five segmentations in order to keep only the variables that were significant. This method was applied using different set of variables, and avoiding problems due to correlation of variables. In the end, one model was calibrated with different combinations of variables for each segmentation approach (Model form A). Two other models were calibrated, one using only curvature (CCR) and AADT (Model form B), and one as base model for each approach, using AADT (Model form B) as the only explanatory variable. For all the models the segment length is included as an offset variable. In Table 3.5 the estimated models are presented with the value of QIC, and standard error and level of significance of variables. Models A, B and C assume, respectively, the following form:

Model A: $$E(Y) = e^{\alpha_{0S} + \alpha_{tS}} \times L \times AADT^{\alpha_{1S}} \times e^{\sum_{1=1}^{n} \beta_{iS} * Var_{iS}} \quad (3.4)$$

Model B: $E(Y) = e^{\alpha_{0S} + \alpha_{tS}} \times L \times AADT^{\alpha_{1S}} \times e^{\sum_{1=1}^{n} \beta_{iS} * CCR}$ (3.5)

Model C: $E(Y) = e^{\alpha_{0S} + \alpha_{tS}} \times L \times AADT^{\alpha_{1S}}$ (3.6)

where:

- E (Y): expected annual crash frequency of random variable Y;
- L: length of road segment [m];
- AADT: average annual daily traffic [veh/day];
- α0S αtS: exponent of constant term of the model, and time trend, where the subscript S indicates the segmentation approach number;
- α1S: exponent of AADT, where the subscript S indicates the segmentation approach number;
- βiS: set of parameters of the regression for different set of variables, with S indicating the segmentation approach number, and i (1=1, 2,…7) the variable;
- VariS: set of variables resulting from the stepwise procedure, for each segmentation approach (S);
- CCR: Curvature change rate [gon/m].

The model calibration results are presented in Table 3.5, while the plots of the cumulative residuals are presented in Figures 3.2-3.4. As is evident from Table 3.5, Segmentation 4, with constant length of segments, allows a greater number of variables to be fit than the others segmentations for the primary Model form (A). The segmentation based only on AADT (Segmentation 3), allows the estimation of a model with five of eight variables considered in the stepwise procedure. Similarly, for the model estimated for Segmentation 2, made up inserting 2 curves and 2 tangents in each segment, five variables were also significant, but the value of QIC is lower than for the Segmentation 3 model. The segmentation that gives the worst results in term of number of variables that could be included in the model is Segmentation 5 in which all variables are constant within each segment. Besides, the model for Segmentation 5 has the highest value of QIC.

For Segmentation 3, which is based on AADT, the variables selected with the stepwise procedure have in general greater standard errors than those selected for other segmentation approaches. This is likely because, in motorways, AADT changes only at interchanges, so the segmentation approach can yield very long segments, with considerable within-segment variation in the other variables that cannot adequately be modeled.

Results reported in Table 3.5, Figure 3.6 and Figure 3.7 show not only differences in the g.o.f. and number of explanatory variables, but also, sometimes, differences in the sign of the coefficients. This indicates, depending on the segmentation, an opposite influence of the variable on the expected number of crashes estimated by the SPF.

In general, Segmentation 2 gives the best results for the primary model form, based on both QIC and the CURE plots. The cumulative residual curves in Figure 3.2 through Figure 3.16

oscillate closer to the value of zero than for the other segmentations.

Table 3.5. Value of regression parameters, (p-value) and [Standard error] for different segmentations (1, 2, 3, 4, 5) and Model form A.

	Seg_1 (Curve Based)	Seg_2 (2 Curves, 2 Tangents)	Seg_3 (AADT Based)	Seg_4 (Fixed Length)	Seg_5 (Homogeneous)
	Multiple variable models from stepwise procedure (Model form A)				
Interc. (α0S+αtS)	-20.4439 (<.0001) [1.0820]	-21.7516 (<.0001) [1.4295]	-13.8951 (0.0003) [3.8420]	-20.1429 (<.0001) [1.3529]	-20.7288 (<.0001) [1.2874]
AADT (α1S)	1.3652 (<.0001) [0.1124]	1.4797 (<.0001) [0.1417]	0.8279 (0.0143) [0.3381]	1.3475 (<.0001) [0.1358]	1.4273 (<.0001) [0.1307]
CCR (βiS)	2508.331 (0.0054) [9.0223]	484.9824 (0.0042) [169.5507]	-2931.75 (0.0273) [1328.017]	262.6808 (0.0066) [96.7806]	0.2111 (0.0003) [0.0585]
I (βiS)	-	-	11.8788 (0.0172) [4.9868]	-14.3209 (<.0001) [3.5159]	-6.0280 (0.0500) [3.1112]
id (βiS)	5.1423 (0.0010) [1.5671]	-	-	-16.2616 (0.0050) [5.7890]	-
Tun(βiS) (Categ)	0.0058 (0.0015) [0.0018]	0.0050 (0.0087) [0.0019]	0.0258 (<.0001) [0.0050]	0.0046 (0.0097) [0.0018]	-0.4540 (<.0001) [0.0981]
RSH6 (βiS)	-	-0.0263 (0.0001) [0.0069]	-0.0634 (0.0004) [0.0178]	-	-
RSH56 (βiS)	-0.0031 (<.0001) [0.0008]	-	-	-0.0037 (<.0001) [0.0009]	-
SCR (βiS)	-	-2.3927 (<.0001) [0.2561]	-	8.4648 (<.0001) [2.10570]	-
QIC	3322.00	1081.65	1761.16	2706.95	4510.73

For model form B, in terms of the CURE plots, the best model is estimated from Segmentation 3, and shown in Figure 2.3.

For model form (C), based only on AADT, the model that oscillates closest to the value of zero is based on Segmentation 4. Only one model exceeds the ±2σ boundary -- the AADT-based model for Segmentation 3 (Figure 4 for the value of AADT close to 15,000 veh/day). Segmentation 5, characterized by constant value of variables inside each segment, gives the poorest results, similar to the earlier conclusion based on QIC.

Table 3.6. Value of regression parameters, (p-value) and [Standard error] for different segmentations (1, 2, 3, 4, 5) and Model form B.

	Model with AADT and CCR (Model form B)				
	Seg_1 (Curve Based)	Seg_2 (2 Curves, 2 Tangents)	Seg_3 (AADT Based)	Seg_4 (Fixed Length)	Seg_5 (Homogeneous)
Interc. (α0S+αtS)	-18.9141 (<.0001) [1.3226]	-20.7128 (<.0001) [1.6767]	-26.3161 (<.0001) [3.4852]	-20.3723 (<.0001) [1.6194]	-19.7873 (<.0001) [1.3327]
AADT (α1S)	1.2075 (<.0001) [0.1364]	1.3713 (<.0001) [0.1667]	1.8705 (<.0001) [0.3207]	1.3476 (<.0001) [0.1613]	1.2891 (<.0001) [0.1358]
CCR (βiS)	23.7961 (0.0021) [7.7267]	489.8783 (0.0031) [165.8877]	2250.728 (0.005) [820.6121]	291.9741 (0.0020) [94.3059]	0.2022 (0.0006) [0.0588]
QIC	3325.57	1109.89	1462.87	2593.60	4580.81

In general, for all models, the CURE plots reveal that the models tend to underestimate the number of crashes for low values of AADT, and to overestimate crashes for higher values of AADT.

The purpose of this study was to investigate the influence of segmentation on the performance of safety performance functions (SPFs), in terms of goodness of fit and the variables that could be modeled. To do this it was necessary to sometimes aggregate variables into a usable form, to have a constant value of each modeled variable in each segment.

Table 3.7. Value of regression parameters, (p-value) and [Standard error] for different segmentations (1, 2, 3, 4, 5) and Model form C.

	Base Model (Model form C)				
	Seg_1 (Curve Based)	**Seg_2 (2 Curves, 2 Tangents)**	**Seg_3 (AADT Based)**	**Seg_4 (Fixed Length)**	**Seg_5 (Homogeneous)**
Interc. (α0S+αtS)	-19.1467 (<.0001) [1.3216]	-18.8182 (<.0001) [1.3281]	-18.2944 (<.0001) [1.1776]	-19.1993 (<.0001) [1.3039]	-18.9870 (<.0001) [1.2908]
AADT (α1S)	1.2358 (<.0001) [0.1353]	1.2000 (<.0001) [0.1363]	1.1515 (<.0001) [0.1208]	1.2403 (<.0001) [0.1328]	1.2163 (<.0001) [0.1321]
QIC	2947.62	1103.22	1363.27	2583.40	4410.09

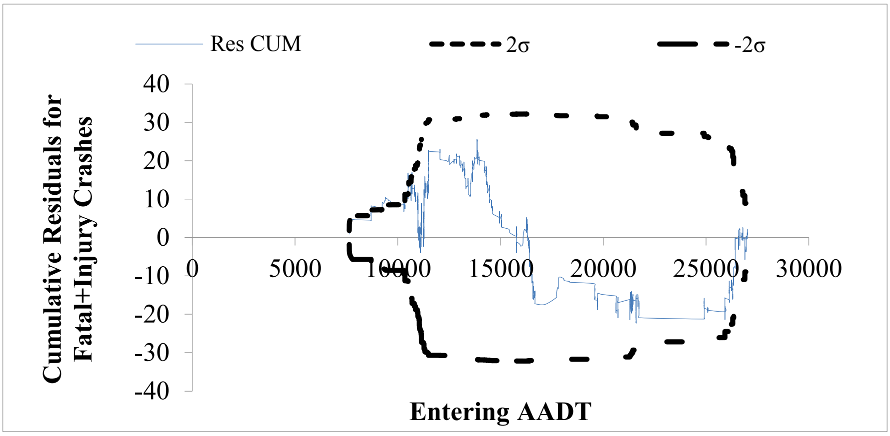


Figure 3.2. CURE Plots with ± 2σ for model form A for segmentation 1.

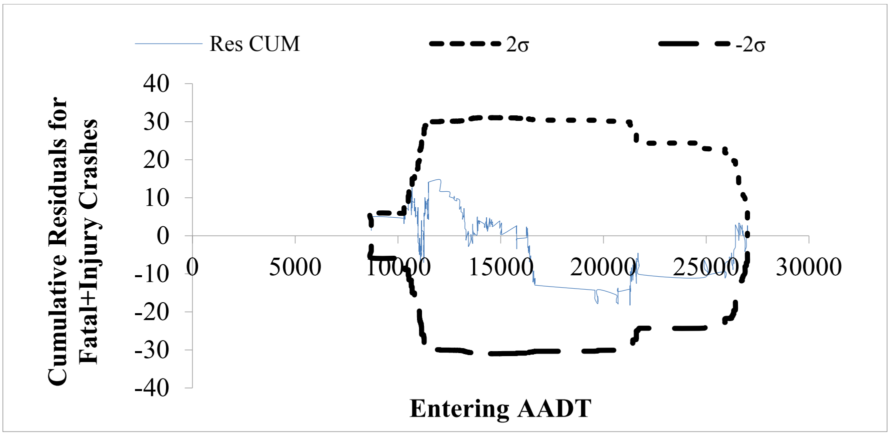


Figure 3.3. CURE Plots with ± 2σ for model form A for segmentation 2.

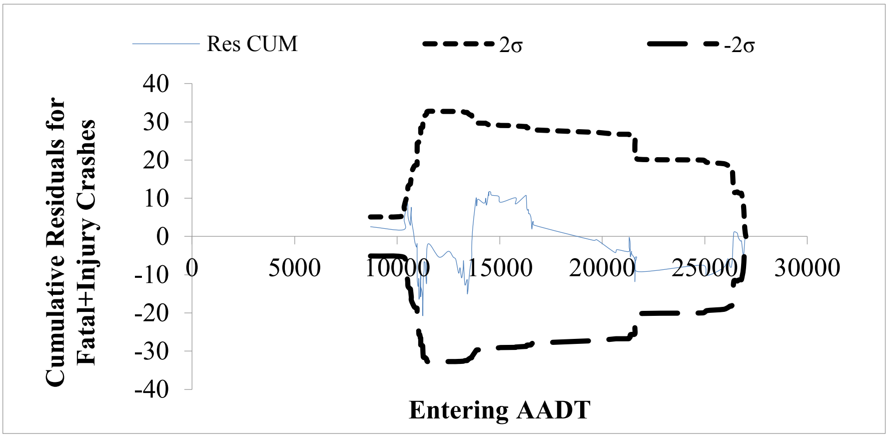


Figure 3.4. CURE Plots with ± 2σ for model form A for segmentation 3.

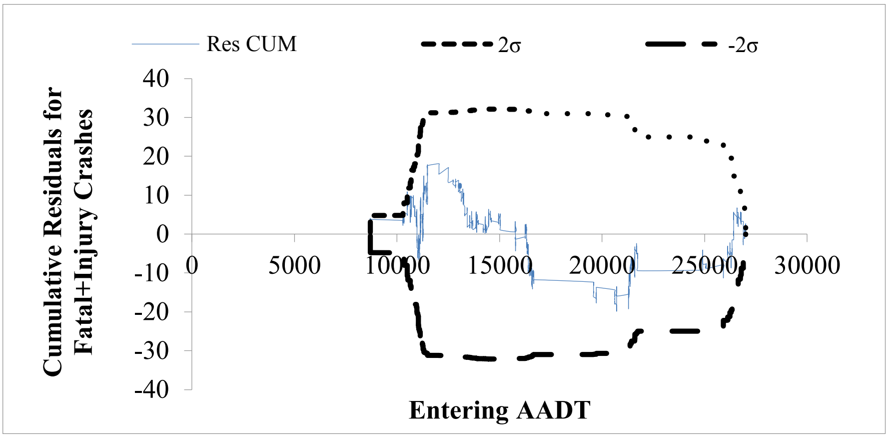


Figure 3.5. CURE Plots with ± 2σ for model form A for segmentation 4.

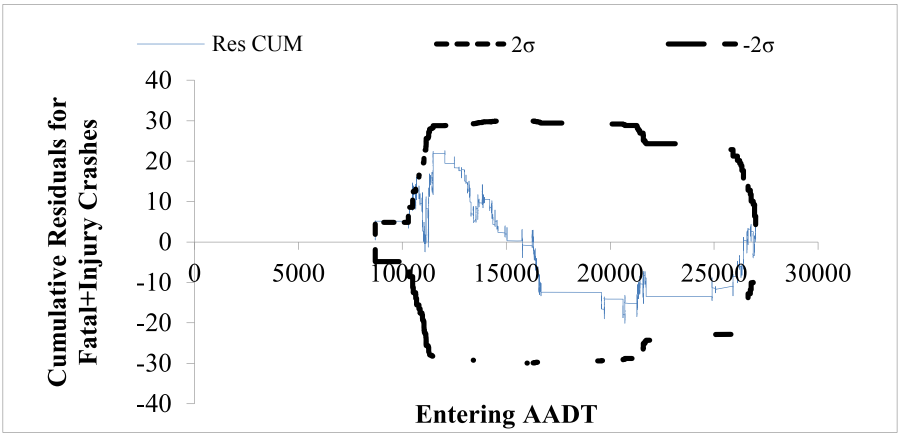


Figure 3.6. CURE Plots with ± 2σ for model form A for segmentation 5.

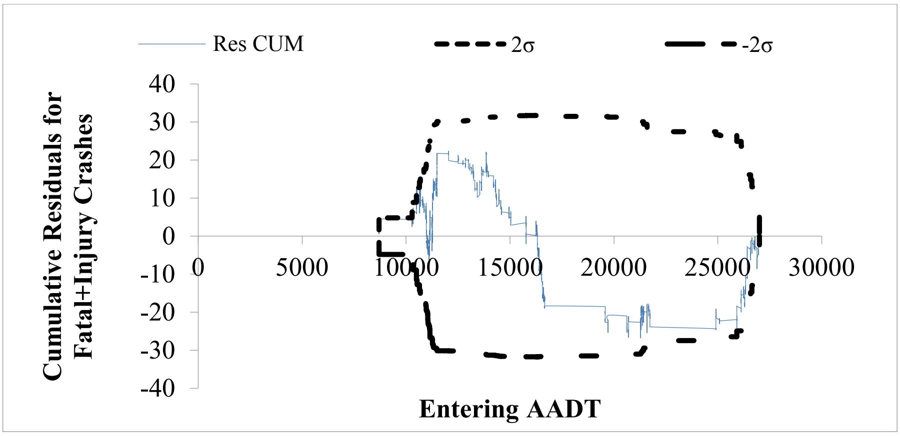


Figure 3.7. CURE Plots with ± 2σ for model form B for segmentation 1.

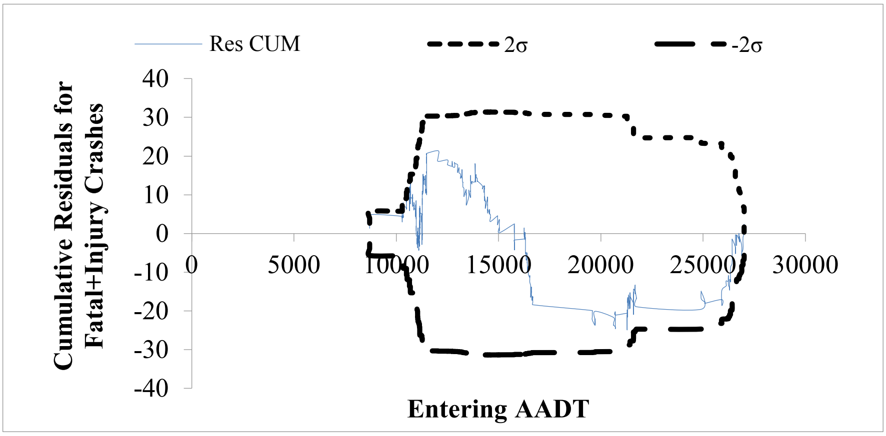


Figure 3.8. CURE Plots with ± 2σ for model form B for segmentation 2.

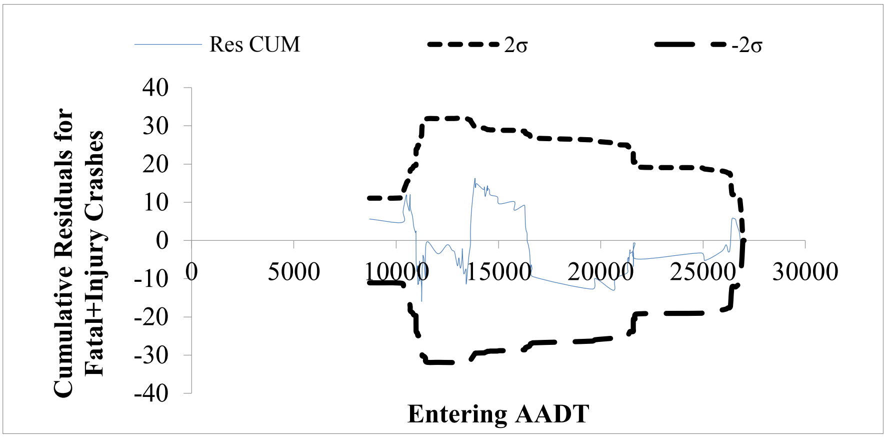


Figure 3.9. CURE Plots with ± 2σ for model form B for segmentation 3.

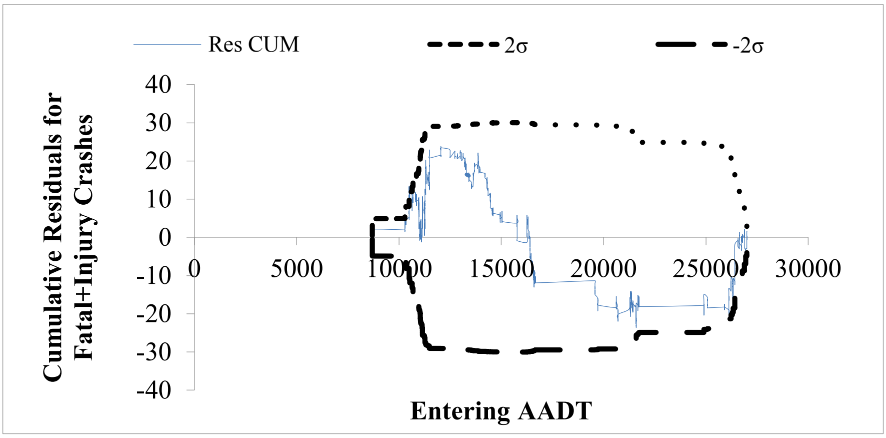


Figure 3.10. CURE Plots with ± 2σ for model form B for segmentation 4.

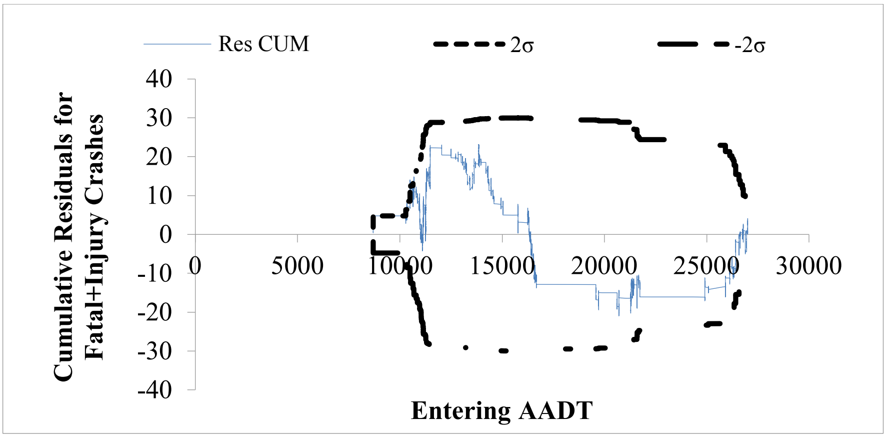


Figure 3.11. CURE Plots with ± 2σ for model form B for segmentation 5.

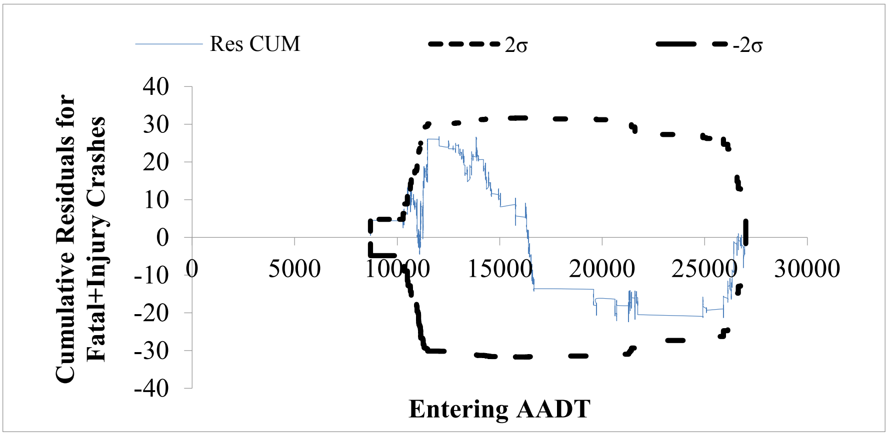


Figure 3.12. CURE Plots with ± 2σ for model form C for segmentation 1.

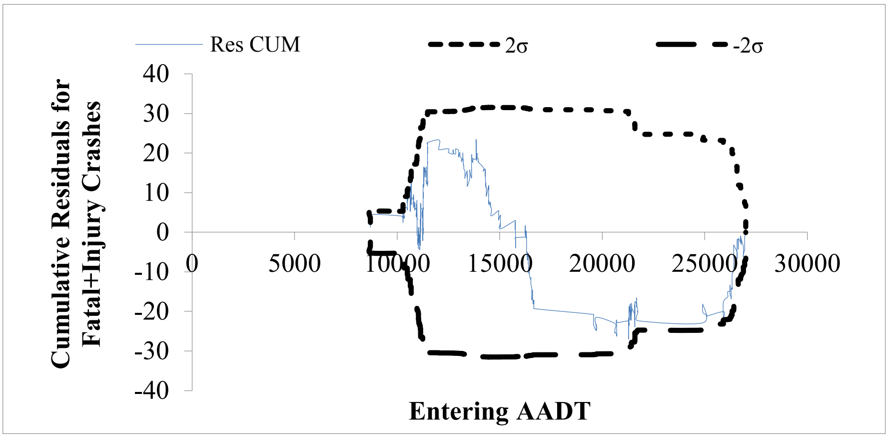


Figure 3.13. CURE Plots with ± 2σ for model form C for segmentation 2.

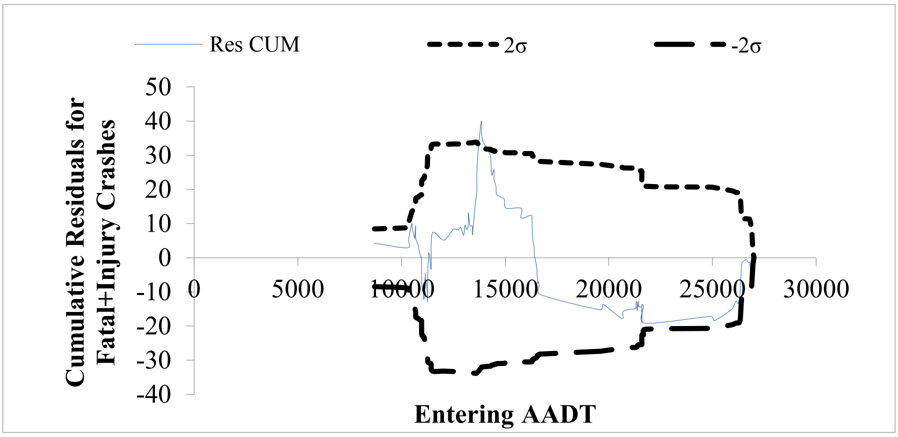


Figure 3.14. CURE Plots with ± 2σ for model form C for segmentation 3.

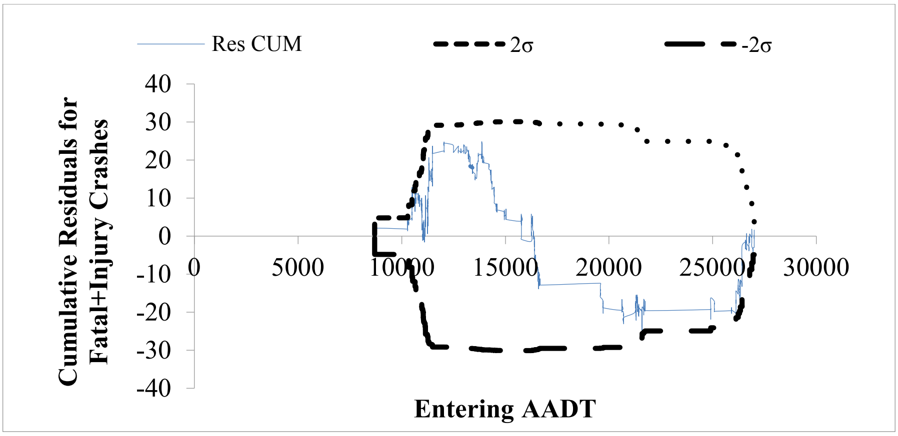


Figure 3.15. CURE Plots with ± 2σ for model form C for segmentation 4.

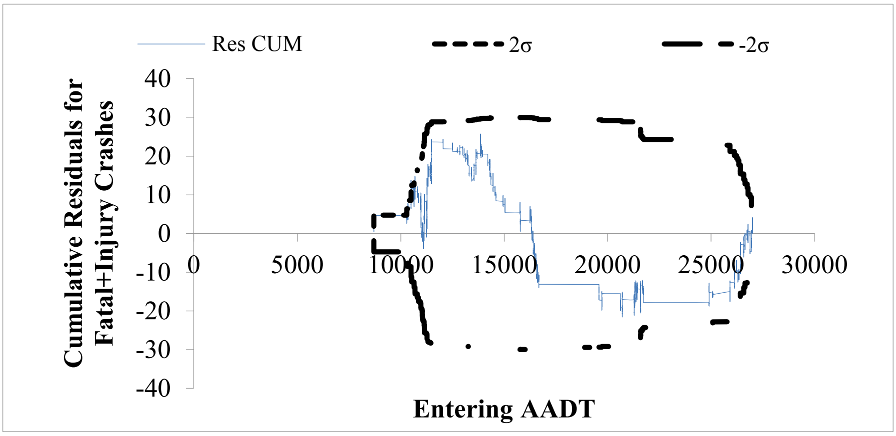


Figure 3.16. CURE Plots with ± 2σ for model form C for segmentation 5.

3.4.1. Summary on the segmentation approach

Five different segmentation approaches were evaluated with three different model forms. One approach is based on the Highway Safety Manual (HSM) method, using curvature and AADT as the basis, one has two curves and two tangents inside each segment, one is based on a constant AADT in each segment, one has a fixed length of each segment, and one has all the variables constant within each segment relative to their original value.

One model was calibrated with different combinations of variables for each segmentation approach. Two other models were calibrated for each approach, one using only curvature and AADT, and one a base model, using AADT as the only explanatory variable. The models were estimated for a sample of rural motorways segments in Italy, using data for the years 2002 through 2009. The Generalized Estimating Equation (GEE) procedure was applied to develop the SPFs, which were evaluated using cumulative residual (CURE) plots and the Quasilikelihood under the Independence model Criterion (QIC) value.

The best results were obtained for the segmentation based on two curves and two tangents (Segmentation 2) and the segmentation with fixed length (Segmentation 4). Segmentation 5, characterized by a constant value of all original variables inside each segment, was the poorest approach by all measures. This is likely because it yields very short segments, resulting in non-perfect identification of the location of crashes and in a large number of segments with zero crashes. Both factors create difficulties in making sound statistical inference.

Fixed length segmentation may be the most flexible in practical applications. This is because the segment length can be determined by data availability and quality, and other factors to optimize the SPF calibration. The length chosen for this research was based on pragmatic reasoning. Given the promise shown by the results, further research can explore alternative considerations for determining the length of fixed length segments used for SPF development. Similarly, the segmentation approach (Segmentation 2) based on two curves and two tangents in each segment, which also showed promise, could be further explored by considering different numbers of curves and tangents in a segment.

3.5. Ranking of the road network

Identification of hazardous locations is the starting point of the process by which locations are selected for safety improvement. Typically, the safety record of a location, along with other information, is used to identify and rank sites that should be investigated for safety deficiencies and possible treatment of these deficiencies. The process is sometimes known as "blackspot identification". More recently, the term "identification of sites with promise" has been used.

All the effort in the development of the techniques for the SPF calibration and to improve their goodness of fit have the target to make the identification of hazardous location as realistic as possible. In the European regulation (96/2008 CE) on Road Safety as well as the Italian transposition of the Code, ranking the sites based on their safety level is a central element in the task required to the Agencies.

3.5.1. Empirical Bayes Ranking

The empirical Bayes analysis introduced by Hauer [3.3] (cfr. Chapter 1) reduce drastically the RTM bias, and represents the state of the art for reliability when the hazardous sites has to be identified. Unfortunately the analysis conducted earlier in the Chapter don't allow the application of the EB correction. Indeed the general estimating equation is a quasi-likelihood method of calibration, and the method doesn't provide parameter able to describe the overdispersion of the data. To overcome this problem the same model were calibrating using a GLM approach and the overdispersion parameter was used to apply the EB correction.

The ranking is performed using the constant length segmentation to avoid that the functional relationship between the overdispersion parameter and the length of segment could influence the estimation [3.23] [3.24]. Moreover using two different techniques, the GEE for the calibration of the SPF, and the GLM to investigate the overdispersion the constant length segmentation can reduce the potential bias in the estimation.

The period of analysis chosen for the ranking are the last three years used in the previous elaboration (2007, 2008 and 2009). This is why longer period can introduce bias due to some changes or treatment applied by the Agency. The multi-variable model form (Model form A) was chosen with the same meaning of symbols:

$$E(Y) = e^{\alpha_0+\alpha_t} \times L \times AADT^{\alpha_1} \times e^{\beta_1 * CCR + \beta_2 * I + \beta_3 * I_d + \beta_4 * Tun + \beta_5 * RSH56 + \beta_6 * SCR} \qquad (3.7)$$

The total number of segments are 226 on the two directions and in Table 2.8 about the 10% is reported. As is it clear from Table

2.8 an heterogeneous value of traffic is present in top ten and the first site of the rank has not the higher value of the predicted crashes.

Moreover comparing the observed reported in the first column of the Table 3.8 with the expected is evident that the observed number of crashes cannot be able to identify sites with safety problem.

Table 3.8. Ranking (10 % of the total segment analyzed) on fixed length segmentation using the empirical Bayes crashes estimation.

Observed	Predicted	Expected (EB)	AADT (Veh/day)	w	Milepost (Starting)
8	3,06	5,70	8696,43	0,47	75.40 R
7	3,79	5,68	12499,85	0,41	40.95 R
6	3,60	4,98	24904,36	0,43	71.50 L
7	2,64	4,81	8696,43	0,50	74.10 R
7	2,36	4,54	19579,09	0,53	63.70 L
5	3,25	4,21	24904,36	0,45	72.15 L
5	2,99	4,05	19579,09	0,47	66.95 L
4	3,61	3,83	8696,43	0,43	74.75 R
4	3,07	3,57	24904,36	0,47	73.45 L
5	2,21	3,48	19717,02	0,55	66.95 R
6	1,70	3,38	15034,70	0,72	69.15 R
4	2,65	3,32	19717,02	0,55	68.90 R
5	1,96	3,25	19579,09	0,56	62.40 L
4	2,52	3,24	24904,36	0,50	70.85 L
3	3,25	3,11	8696,43	0,47	70.20 L
3	3,19	3,09	24904,36	0,58	76.05 L
8	1,07	3,06	10496,21	0,57	18.85 L
8	1,05	3,00	10288,58	0,76	37.75 L
4	2,07	2,92	12499,85	0,71	39.65 R
4	2,05	2,89	19717,02	0,63	67.60 R
4	1,99	2,85	14947,66	0,66	52.00 L
2	3,53	2,66	24904,36	0,51	74.75 L
2	3,40	2,62	8696,43	0,57	76.05 R
k= 0.3744 - Segment Length = 650 m					

3.5.2. *Potential for Safety Improvement (PSI) analysis*

Persaud et al. [3.25], among others, have proposed that sites be ranked according to their potential for safety improvement (S), to be estimated as the difference between the expected number of crashes (m) and what is normal for similar sites,

$$S = (m - P) \tag{3.8}$$

Where P is an estimate of the predicted (by an SPF) annual number of crashes and m the expected number of crashes (Empirical Bayes Correction). In particularly m is estimated inserting in the model all the significant variable, P is related to the base model (only AADT).

The basis of this index was first suggested by McGuigan [3.26], who called it the potential accident reduction, and by Jorgensen [3.27]. However, they had suggested using the accident count instead of m in Equation 2.8, an application that would create difficulties due to random fluctuation in counts where, as is often the case, a relatively short accident history is used. More recently, Tarko et al. [3.28] sought to overcome this difficulty by using the concept of the confidence level for which the above-norm number of crashes is larger than zero.

A validation of the potential-for-safety-improvement concept [3.25] showed that the method is not only conceptually sound but is also comparatively efficient.

However Persaud et al [3.29] suggested an update methodology for the application of the PSI. In particularly two practical options are suggested:

Option 1

$$S = (m - P_T) \quad (3.9)$$

Here m is still calculated by the best possible regression models, incorporating available variables that may contribute to unsafety, and P_T is based on a model that includes traffic volume but no treatable variables.

Option 2

$$S = (m - P_B) \quad (3.10)$$

Here the difference is that P_B is for a base condition, reflecting the fact that what is normal can be found in the predominant values of the treatable variables. The idea is that roads are usually built to some desirable standard from a safety point of view and that the standard values will predominate in a data set and will reflect a desirable or achievable level of safety.

The results of the application of the method are reported in Table 3.9 calculated using the option 1 and the model form of Equation 2.7 for m while P is calculated with the base model.

As it clear from Table 3.9, the ranking based on PSI gives different results than the expected. The differences are due to the combination of the predicted and the expected in the computation of the index. Considering "normal" for sites with similar characteristic the expected any drift due to the multi-variable model could be identified by the PSI index. As such it represent one of the most reliable index for the identification of hazardous location.

Table 3.9. Ranking (10 % of the total segment analyzed) on fixed length segmentation using the Potential for Safety Improvement Index (PSI).

Observed	Predicted	Expected (EB)	AADT (Veh/day)	PSI	Milepost (Starting)
7	0,54	5,68	12499,85	5,14	40.95 R
8	1,23	5,70	8696,43	4,47	75.40 R
6	1,26	4,98	24904,36	3,72	71.50 L
7	1,23	4,81	8696,43	3,58	74.10 R
7	0,97	4,54	19579,09	3,57	63.70 L
5	0,97	4,05	19579,09	3,08	66.95 L
5	1,26	4,21	24904,36	2,95	72.15 L
6	0,69	3,38	15034,70	2,69	59.15 R
8	0,43	3,06	10496,21	2,63	18,85 L
4	1,23	3,83	8696,43	2,60	66.95 R
8	0,13	3,00	10288,58	2,59	37.75 L
5	0,29	3,48	19717,02	2,52	66.95 R
4	0,16	2,92	12499,85	2,37	39.65 R
4	0,29	3,32	19717,02	2,37	68.90R
4	0,38	3,57	24904,36	2,31	73.45 D
5	0,28	3,25	19579,09	2,28	62.40L
4	0,20	2,85	14947,66	2,15	52.00 L
8	0,13	2,53	10338,63	2,12	37.75 R
6	0,13	2,47	10338,63	2,06	27.95 R
4	0,13	2,46	10496,21	2,03	13.00 L
5	0,17	2,57	13209,75	2,00	48.10 L
4	0,38	3,24	24904,36	1,98	70.85 L
4	0,29	2,89	19717,02	1,94	67.60 R
k= 0.3744 - Segment Length = 650 m					

In Table 3.10 a comparison of the two methodology is reported. In order to evaluate the performance of ranking of both the EB and the PSI, based on this latter a rank on the first 10% of hazardous sites is reported indicating with A the EB approach and with B the PSI index. In the second and third column of the Table 3.10 the ranking position for criteria A and B is reported respectively. Considering the 10% of the entire analyzed path inly 3 segments on 23 are not identify as hazardous by the EB methodology while are included in the PSI index ranking.

Table 3.10. Comparison of the two ranking performed (PSI and EB) with reference on the PSI index.

Observed	PSI - Rank	EB - Rank	PSI	Expected (EB)	Milepost (Starting)
7	1B	2A	5,14	5,68	40.95 R
8	2B	1A	4,47	5,70	75.40 R
6	3B	3A	3,72	4,98	71.50 L
7	4B	4A	3,58	4,81	74.10 R
7	5B	5A	3,57	4,54	63.70 L
5	6B	7A	3,08	4,05	66.95 L
5	7B	6A	2,95	4,21	72.15 L
6	8B	11A	2,69	3,38	59.15 R
8	9B	17A	2,63	3,06	18,85 L
4	10B	8A	2,60	3,83	66.95 R
8	11B	18A	2,59	3,00	37.75 L
5	12B	10A	2,52	3,48	66.95 R
4	13B	19A	2,37	2,92	39.65 R
4	14B	12A	2,37	3,32	68.90R
4	15B	9A	2,31	3,57	73.45 D
5	16B	13A	2,28	3,25	62.40L
4	17B	21A	2,15	2,85	52.00 L
8	18B	-	2,12	-	37.75 R
6	19B	-	2,06	-	27.95 R
4	20B	-	2,03	-	13.00 L
5	21B	-	2,00	-	48.10 L
4	22B	14A	1,98	3,24	70.85 L
4	23B	20A	1,94	2,89	67.60 R
k= 0.3744 - Segment Length = 650 m					

3.6. Chapter summary

In the present Chapter 3 a methodology for the optimization of the goodness of fit of the model, varying the segmentation approach is reported based on a study by Cafiso et al. presented at the 92nd Transportation Research Board in 2012 [3.1]. Five different segmentation approach were tested on three different model form, with the results that longer segments give the best performance in term of reliability. The QIC, together with the CURE plot methodology were the gof measure used in the analysis. Particularly the fixed length segmentation approach give the best results in term of number of variables could be included in the model and goodness of fit performance together with the segmentation which includes 2 curves and two tangents inside each segment. The goodness of fit of the model is one of the main issue where the models are used for the identification of hazardous location.

For the safety site ranking the EB methodology was applied on the fixed length segmentation. To estimate the overdispersion parameter the model was recalibrated using the GLM approach. The top 10% of total number of segment was reported showing that the ranking based on the observed number of crashes gives different results. The ranking analysis was then compared with the Potential for Safety Improvement Index showing that the two methodology give almost the same results. Particularly on a total amount of 23 segments (about the 10% of the total), based on the

PSI ranking, only 4 segments are not identified as hazardous by the EB ranking methodology.

References

[3.1]. Cafiso. S.. D'Agostino. C.. Persaud. B.. Investigating the influence of segmentation in estimating safety performance functions for roadway sections. Transportation Research Board Annual Meeting Proceedings. Washington D.C.. 2013.

[3.2]. Highway Safety Manual (HSM), First Edition. FHWA American Association of State Highway and Transportation Officials, U.S.A, 2010.

[3.3]. Hauer, E.. Observational Before-After Studies in Road Safety. Pergamon Press, Elsevier Science Ltd., Oxford, England, 1997.

[3.4]. Resende, P., Benekohal, R.. Effect of roadway section length on accident modeling traffic congestion and traffic safety. In: The 21st Century Conference, ASCE, Chicago, IL, 1997.

[3.5]. Quin, X., Wellner, A., Segment length impact on highway safety screening analysis. TRB 2012 Annual Meeting, Washington, D.C., 2012.

[3.6]. Miaou, S-P., and Lum, H. Modeling Vehicle Accidents and Highway Geometric Design Relationships. Accident Analysis and Prevention, Vol. 25, No. 6, pp. 689-709, 1993.

[3.7]. Ogle, J., Alluri, P., and Sarasua, W. MMUCC and MIRE: The role of segmentation in safety analysis. Transportation

Research Board Annual Meeting Proceedings, Washington D.C., 2011.

[3.8]. Cafiso S., Di Silvestro G.. Performance of Safety Indicators in Identification of Black Spots on Two-Lane Rural Roads. Transportation research Record, n. 2237, Safety, Data, Analysis and Evaluation,Vol.2, pp. 78-87, 2011.

[3.9]. Cenek, P. D., R. B. Davies, M. W. McLarin, G. Griffith-Jones, and N. J. Locke. Road Environment and Traffic Crashes. Transfund New Zealand Research Report 79. Wellington, 1997.

[3.10]. Cafiso S., A. Di Graziano, G. Di Silvestro, and G. La Cava. Safety Performance Indicators for Local Rural Roads: Comprehensive Procedure from Low-Cost Data Survey to Accident Prediction Model. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.

[3.11]. Koorey G.. Road Data Aggregation and Sectioning Considerations for Crash Analysis. Transportation Research Record: Journal of the Transportation Research Board, No. 2103, Transportation Research Board of the National Academies, Washington, D.C, pp. 61–68, 2009.

[3.12]. I dati sugli incidenti stradali rilevati nel 2002, 2003, 2004, 2005, 2006 2007, 2008 e 2009. Istituto Nazionale di Statistica, Milan, Italy, 2010.

[3.13]. Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., Persaud, B.. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. Accident Analysis and Prevention, Vol. 42 pp. 1072–1079, 2010.

[3.14]. Cafiso, S., La Cava, G., Montella, A., Pappalardo, G.. A procedure to improve safety inspections effectiveness and reliability on rural two-lane highways. The Baltic Journal of Road and Bridge Engineering 1 (3), 143–150, 2006.

[3.15]. Cafiso, S., La Cava, G., Montella, A.,. Safety index for evaluation of two-lane rural highways. Transportation Research Record 2019, 136–145, 2007.

[3.16]. Cafiso, S., La Cava, G., Montella, A., Di Graziano, A., Pappalardo, G., Taormina, S.,. Final research report. Identification of hazard location and ranking of measures to improve safety on local rural roads: EU IASP project. http://ec.europa.eu/transport/roadsafety/publications/projectfiles/iaspen.htm, 2007.

[3.17]. Cafiso, S., La Cava, G., Montella, A., Pappalardo, G.. Operative procedures for safety inspections on two-lane rural roads. In: ARACNE, Rome, Italy, 2008.

[3.18]. Lord, D. and Persaud, B.. Accident Prediction Models With and Without Trend. Transportation Research Record 1717 Paper No. 00-0496, 2007.

[3.19]. Pan, W.. Akaike's information criterion in generalized estimating equations. Biometrics, 57, 120-125, 2001.

[3.20]. Hardin, J.W. and Hilbe, J.M.. Generalized Estimating Equations. Chapman & Hall/CRC: New York, 2003.

[3.21]. Hauer, E., and J. Bamfo. Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables. Proc., ICTCT Conference, Lund, Sweden, 1997.

[3.22]. SAS Institute Inc. SAS Help and Documentation: QIC, QICu. Software of SAS 9.2, Copyright (c) 2002-2008, Cary, NC, USA. http://support.sas.com/kb/26/100.html. Accessed July 10, 2012.

[3.23]. Hauer. E.. Overdispersion in Modelling Accidents on Road Sections and in Empirical Bayes Estimation. Accident Analysis and Prevention. Vol. 33. No. 6. 2001. pp. 799–808.

[3.24]. Cafiso. S.. Di Silvestro. Persaud. B.. Begum. M.A.. Revisiting Variability of Dispersion Parameter of Safety Performance for Two-Lane Rural Roads. Transportation Research Record: Journal of the Transportation Research Board. No. 2148. Transportation Research Board of the National Academies. Washington. D.C.. 2010. pp. 38–46.

[3.25]. Persaud, B. N., W. Cook, and A. Kazakov. Demonstration of New Approaches for Identifying Hazardous Locations and Prioritizing Safety Treatment In Proc., 7th International

Conference: Traffic Safety on Two Continents, Lisbon, Portugal, September 22–24, 1997.

[3.26]. McGuigan, D. R. D. The Use of Relationships Between Road Accidents and Traffic Flow in ‘Black-Spot’ Identification. Traffic Engineering and Control, Aug.–Sept. 1981, pp. 448–453.

[3.27]. Jorgensen, N. O. Statistical Detection of Accident Blackspots. OTAPIARC 11th International Study Week in Transportation and Safety.Brussels, 1972.

[3.28]. Tarko, A., K. Sinha, and O. Farooq. Methodology for Identifying Highway Safety Problem Areas. In Transportation Research Record 1542, TRB, National Research Council, Washington, D.C., 1996, pp. 49–53.

[3.29]. Persaud, B. N., Lyon G., Nguyen, T.. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. Transportation Research Record 1665. Paper No. 99-0534. 1999.

CHAPTER 4

CRASH MODIFICATION FACTOR AND FUNCTION

4.1. Introduction

There is a growing attention to road safety, in Europe and Italy as well. New regulations applied on TERN network (motorways in Italy) push Agencies to introduce new methodological approaches to Road Safety, monitoring the treatment and controlling the level of safety on the managed road network. The application of the HSM doesn't always provide adequate results. The problem related to transferability of the SPFs is a clear example of how the model developed in other Countries are not always able to catch the safety level of different infrastructures. The key point is that quantification of the expected reduction of crashes related to different treatments, can affect choices and plays a fundamental role in the decision making process.

Together with SPFs the Highway Safety Manual [4.1] introduced the "Crash Modification Factor" (CMF).

A Crash Modification Factor is a multiplicative factor used to compute the expected number of crashes after implementing a given countermeasure at a specific site. The CMF is multiplied by the expected crash frequency without treatment. A CMF greater than 1.0 indicates an expected increase in crashes, while a value less than 1.0 indicates an expected reduction in crashes after implementation of a given countermeasure.

CMFs provide a general idea of the safety effects of a countermeasure, indeed, they could improve the accuracy of the choices introducing the new concept through each treatment has a different impact on safety above all on the crash category directly related to the treatment.

To calibrate CMFs, it is necessary to know how many crashes are expected without countermeasures. Specifically, the annual expected number of crashes without treatment is multiplied by the CMF to estimate the expected number of crashes with treatment. Estimating the expected crashes without treatment is not a trivial task; it is not simply the number of observed crashes before treatment, since this value could be higher or lower than expected due to regression-to-the-mean. Also, changes in traffic volume will cause changes in expected crashes.

To develop a CMF, fundamentally, two ways are preferred. It is possible to estimate the CMF from the coefficient of the model, or as a constant value from an analysis of comparison made with a controlled experiments or observational before-after studies or with a cross sectional analysis where the latter is not feasible [4.1] [4.2]. The basic idea is to identify, in the most reliable way, the safety performance of the treatment and quantify the reduction in frequency and/or severity of crashes.

However it is not always reasonable to assume a uniform safety effect for all sites with different characteristics (e.g., safety benefits may be greater for sites with high traffic volumes) [4.3]. A countermeasure may also have several levels or potential values, for this reason HSM, and others studies are developing Crash Modification Function. A CMF is a formula used to compute the CMF for a specific site based on its characteristics. A crash modification function allows the CMF to change over the range of a variable or combination of variables.

In the present chapter the techniques of calibration of Crash Modification Factors are described. At the end of the chapter a CMF for new European Standard of barriers is reported as a case study using the before-after empirical Bayes methodology. That study, by Cafiso et al. [4.5] is in press in the TRB 93rd Annual Meeting and a great part of the references which contributing in the Chapter. As integration of that study a validation of the dataset used for the elaboration is performed and a Crash Modification Function is calibrated varying the curvature for the ran-off-road crashes.

4.2. Methodology

It is important to note that a CMF represents the long-term expected change in crash frequency. Also, a CMF may be based on the crash experience at a limited number of study sites. As such, the actual change in crashes observed after treatment will vary by location and by year.

Usually the data used in the study designs can typically be classified as either before-after or cross-sectional.

Before-after designs include a treatment at some period in time and a comparison of the safety performance before and after treatment for a site or group of sites. Cross-sectional designs compare the safety performance of a site or group of sites with the treatment of interest to similar sites without the treatment at a single point in time. Both before-after and cross-sectional study designs have issues that need to be considered in the development and application of CMFs that are discussed in more detail later.

The knowledge and great part of the reference, related to the method of CMF calibration are taken from a document titled "A guide to developing quality Crash Modification Factor" of Federal Highway Administration and sponsorship by the U.S. Department of Transportation [4.3], in which together with the cross sectional and before-after analysis, reported below, are present others less used techniques for CMF calibration. Together with the method of calibration some significant studies are reported.

CMFs derived from before-after data are based on the change in safety performance due to the implementation of some treatment. There are two fundamental issues with deriving quality CMFs from before-after designs.

1) Sample Size: The required sample size depends on the magnitude of the treatment effect and the uncertainty of the estimate (i.e., the standard error). Generally, the standard error decreases as the sample size increases. As such, one can reduce the uncertainty of an estimate by increasing the sample size.
2) Potential Bias: The observed change in crash experience at treated sites between the periods before and after treatment may be due not only to the countermeasure, but to other factors as well. Other factors include:
    a. Traffic volume changes.
    b. Changes in reported crash experience.
    c. Regression-to-the-mean.

Simple before-after comparisons, also known as naïve before-after studies [4.4], do not account for these changes. As a result, CMFs derived from such studies are usually considered unreliable and rated as being of poor quality. In the before after studies, as mentioned above, the main issue is to find the expected crash frequency in the treated sites if the treatment would not been applied. Based on the method of investigation of this frequency the before-after study can be divided in:

- Before-after studies with comparison group;
- Before-after empirical Bayes analysis;
- Before-after Full Bayes analysis.

The methodology for calibrating the CMF is the same in the methods listed above, what really change is the methodology of estimation of the expected crashes in the after period if the treatment would not have been applied.

For that reason only the Before-after empirical Bayes analysis is described in detail below, the other methods only pro and cons are reported. Moreover, at the end of the paragraph some study design of particular interest are presented.

4.2.1. Before-after with comparison group analysis

A before-after with comparison group study uses an untreated comparison group of sites similar to the treated ones to account for changes in crashes unrelated to the treatment such as time and traffic volume trends. The comparison group is used to calculate the ratio of observed crash frequency in the after period to that in the before period. The observed crash frequency in the before period at a treatment site group is multiplied by this comparison ratio to provide an estimate of expected crashes at the treatment group had no treatment been applied. This is then compared to the observed crashes in the after period at the treatment site group to estimate the safety effects of the treatment.

Ideally, the comparison group should be drawn from the same jurisdiction as the treatment group and be similar to the treatment group in terms of geometric and operational characteristics. The difficulty is that the pool available for the comparison group could be too small if most or all sites are treated or at least affected by the treatment.

This method will not account for regression-to-the-mean unless treatment and comparison sites are also matched on the basis of the observed crash frequency in the before period. Specifically, a control site would need to be matched to each treated site based on the annual crashes in the before period. There are immense practical difficulties in achieving an ideal comparison group to account for regression-to-the-mean (i.e., matching on the basis of crash occurrence) as illustrated in Pendleton et al. [4.6] . In addition, the necessary assumption that the comparison group is unaffected by the treatment is difficult to test and can be an unreasonable assumption in some situations.

Where there is no regression-to-the-mean and where a suitable comparison group is available, the comparison group methodology can be a simple alternative to the more complex empirical Bayes approach. This may be true in cases where 1) crash frequency is not considered in selecting a site for safety treatment, 2) the safety evaluation is strictly related to a change implemented for operational reasons, or 3) a blanket treatment is applied to all sites of a given type. In practice, except for blanket treatments, it is difficult to ascertain that there is no regression-to-the-mean and only a truly random selection of sites for treatment will ensure that there is no selection bias.

A suitable comparison group is one where the ratios of expected crash counts in the after period to the expected crash counts in the before period are equal for the comparison group and the treatment group, had no treatment been applied. For example, if the expected crash count at treated sites were to increase by 10 percent in the after period without treatment then a perfect comparison group should also show this expected increase of 10 percent. Naturally, it is difficult to achieve a perfect comparison group, since the change in crashes at the treatment sites without treatment cannot be known (since there is treatment).

The suitability of a comparison group can be determined by performing a test of comparability for the treatment group and potential comparison groups that was implemented by Hauer [4.3]. The test of comparability compares a time series of target crashes for a treatment group and a candidate comparison group during a period before the treatment is implemented. If a candidate reference group is a good one, then the yearly trends in accident frequencies track each other well over time. As mentioned above, Hauer [4.4] proposes calculating a sequence of sample odds ratios using 1 year of “before” data and the following year as the “after” data, starting with years 1 and 2 and incrementally increasing by 1 year. From this sequence of ratios, the sample mean and standard error is determined. If this sample mean is not sufficiently close to 1.0, and the 95% level of confidence doesn’t include 1, then the candidate reference group is unsuitable.

If the sums of the observed and the expected (Greek letter) for the before period are K and κ for the treatment group; M and μ for the reference group and for the after period L and λ for the treatment group; N and ν for the reference group and are r_c= ν/ μ and r_t= π/κ respectively the ratio of the expected accident counts for the reference and the treatment group and π indicates the sums of expected crashes in the treatment group in the before period if the treatment not be implemented, the ratio:

$$\omega = \frac{r_c}{r_t} \tag{4.1}$$

Is define as the (ω)"odds ratio". The best estimator of ω is given by:

o=(KN)/(LM)/(1+1/L+1/M) (4.2)

At the same way it is possible to estimate the variance of ω as:

VAR{ω}=s^2-(1/K+1/L+1/M+1/N) (4.3)

Where s^2 is the sample variance of o.

An application of the comparison analysis for the reference group is showed below where a Crash Modification Factor is calibrated for different barriers typology.

Additional requirements of a suitable comparison group, as outlined by Hauer [4.4] include:

1) The before and after periods for the treatment and comparison group should be the same.

2) There should be reason to believe that the change in factors other than the treatment under study (e.g., traffic volume changes), which influence safety are the same in the treatment and comparison groups.
3) The crash counts must be sufficiently large.

Indeed in a comparison group before-after safety evaluation, it is vital to ensure that enough crashes are included such that the expected change in safety can be statistically detected. Recall that a statistically significant CMF means that one can say with a given level of significance that the confidence interval for the CMF does not include 1.0.

The four variables that impact whether or not a sample is sufficiently large are:

1) The size of the treatment group, in terms of the number of crashes in the before period.
2) The relative duration of the before and after periods.
3) The likely (postulated) CMF value.
4) The size of the comparison group in terms of the number of crashes in the before and after periods.

It is challenging to assess the adequacy of a sample before collecting data because it is necessary to estimate the number of crashes in the sample that is yet to be collected and develop an intelligent guess about the magnitude of the CMF. These variables impact the precision (standard error) with which the CMF is estimated. A detailed explanation of sample size considerations, as well as estimation methods, are reported by Hauer [4.4].

In the method of comparison group before-after analysis if x is the observed number of crashes in the before period for the treatment group and c is the comparison ratio, or better the ratio between the observed in the reference group in the after (d) and before (p) period, the expected number of crashes for the treatment group that would have occurred in the after period without treatment (B) is estimated from the following equation:

$$B = x \cdot \frac{d}{p} = x \cdot c \qquad (4.4)$$

If the comparison group is suitable, that is, if the crash trends in that group and the treatment group are similar as determined by the test for comparability, the variance of B is estimated approximately the following equation:

$$Var\{B\} = B^2 \cdot \left(\frac{1}{x} + \frac{1}{p} + \frac{1}{d}\right) \qquad (4.5)$$

This estimate is only an approximation since it applies to an ideal comparison group with yearly trends identical to the treatment group, a situation that is practically impossible. A more precise estimate can be obtained by applying a modification, which is typically minor, as derived in Hauer [4.4]. Estimating this modification is not trivial, so it is recommended to estimate the variance assuming an ideal comparison group and recognize that this estimate is a conservatively low approximation. In the ideal case, the CMF and its variance are estimated from Equation 3.7 and 3.8 considering *A* as the crash count in the after period:

$$CMF = \frac{A}{B} \cdot \left(1 + \frac{Var\{B\}}{B^2}\right) \tag{4.6}$$

$$Var\{CMF\} = CMF^2 \cdot \frac{\left(\frac{1}{A} + \frac{Var\{B\}}{B^2}\right)}{\left(1 + \frac{Var\{B\}}{B^2}\right)^2} \tag{4.7}$$

4.2.2. <u>Empirical Bayes before-after analysis</u>

The Empirical Bayes (EB) method is considered the state of the art evaluation procedure developed to perform a before after analysis with the aim to calibrate a CMF. This approach can address the regression-to-the-mean phenomena and properly account for changes in traffic volumes and other variables expressly controlled. The EB method uses two clues in order to estimate the crash count of an entity without treatment:

- the crash records of that entity; and
- the crash frequency expected at similar entities.

In the EB method, a safety performance function is used to estimate the expected crash frequency at the treated locations had modifications not been made. The method has been pioneered by Hauer [4.4] and used by many others in recent evaluations [4.7] [4.8] [4.9] [4.10] [4.11]. The observed counts and the expected counts are combined to produce an improved estimate of the crash frequency in the after period. A simple representation of the technique is shown in Figure 4.1 [4.8]. The modeled or expected crash frequency is combined with the observed crash frequency to produce the adjusted estimate. The weight given to each value is dependent on the years of data used and the overdispersion parameter estimate from the regression models. The more years of data in the observed crash counts the more “weight” that observation is given in combining the two results.

The objective of the empirical Bayes methodology is to more precisely estimate the number of crashes (Previously denoted as B) that would have occurred at an individual treated site in the after period had a treatment not been implemented. Similar to the comparison group method, the effect of the safety treatment is estimated by comparing the sum of the estimates of B for all treated sites with the number of crashes actually recorded after treatment.

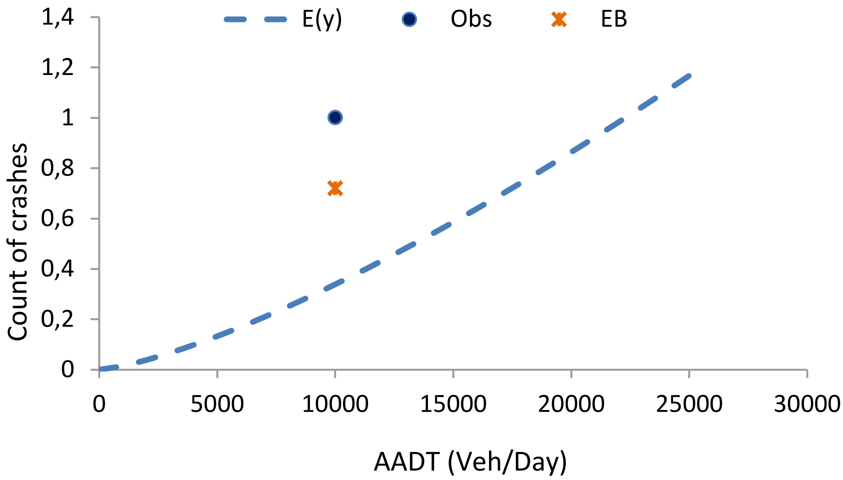


Figure 4.1. Representation of EB Estimate.

The advantage of the empirical Bayes approach is that it correctly accounts for observed changes in crash frequencies before and after a treatment that may be due to regression-to-the-mean. In doing so, it also facilitates a better approach than the comparison group method for accounting for changes in safety due to traffic volumes and time trends.

As mentioned above in the empirical Bayes evaluation of the effect of a treatment, the change in safety for a given crash type, with the same meaning of symbols, is given by:

$$B - A \quad (4.8)$$

Because of changes in safety that may result from changes in traffic volume, from regression-to-the-mean, and from trends in crash reporting and other factors, the count of crashes before a treatment by itself is not a good estimate of B [4.12] a reality that has now gained common acceptance. Instead, B is estimated from an empirical Bayes (EB) procedure in which a safety performance function (SPF) is used to first estimate the number of crashes that would be expected in each year of the "before" period at locations with traffic volumes and other characteristics similar to a treatment site being analyzed. The sum of these annual SPF estimates (P) is then combined with the count of crashes (x) in the before period at the treatment site to obtain an estimate of the expected number of crashes (m) before the treatment. This estimate of m is

$$m = w1(x) + w2(P) \quad (4.9)$$

The weights w1 and w2 are estimated as

$$w_1 = \frac{P}{P + \frac{1}{k}} \quad (4.10)$$

$$w_2 = \frac{1}{k\left(P + \frac{1}{k}\right)} \quad (4.11)$$

where k is the dispersion parameter of the negative binomial distribution that is assumed for the crash counts used in estimating the SPF. The value of k is estimated from the SPF calibration process with the use of a maximum likelihood procedure.

The comparison ratio is then applied to m from Eq. (4.9) to account for the length of the after period as well as differences in traffic volumes and general trends in crash risk due to factors such as weather, reporting practices and the other safety countermeasures between the before and after periods. The result, after applying this factor, is an estimate of B. The procedure also produces an estimate of the variance of B, the expected number of crashes that would have occurred in the after period without the treatment.

The estimate of B is then summed over all road sections in a treatment group of interest (to obtain B_{sum}) and compared with the count of crashes during the after period in that group (A_{sum}). Taking into account that the significance of the difference (B-A) is established from this estimate of the variance of B, calculated as follow:

$$Var(B_{sum}) = C \cdot m \cdot w_2 \quad (4.12)$$

and assuming, on the basis of a Poisson distribution of counts, that $Var\{A\} = A$. The variance of B is also summed over all sections in the group of interest.

The CMF called also index of safety effectiveness (θ) is estimated as:

$$CMF = \vartheta = \frac{A_{sum} / B_{sum}}{1 + \left[Var(B_{sum}) / B_{sum}^2\right]} \tag{4.13}$$

The standard deviation of θ is given by:

$$Stddev(\vartheta) = \left[\frac{\vartheta^2 \left\{ \left[Var(A_{sum}) / A_{sum}^2\right] + \left[Var(B_{sum}) / B_{sum}^2\right] \right\}}{\left[1 + Var(B_{Sum}) / B_{sum}^2\right]^2} \right]^{0.5} \tag{4.14}$$

When an empirical Bayes before-after safety evaluation have to be implemented, it is vital to ensure that enough data are included such that the expected change in safety can be statistically detected. Currently, there is no formal method for determining required sample sizes for the empirical Bayes before-after approach. The method presented in Hauer [4.3] pertains to the comparison group method and can be used to approximate the sample size required for an empirical Bayes study. The sample size estimates could be considered conservative in that the empirical Bayes approach reduces uncertainty in the estimate of expected crashes.

The observed change in crash experience at treated sites between the periods before and after treatment may be due not only to the countermeasure, but to other factors as well. If these factors are not properly accounted for, there is the potential to bias the results. These other factors include:

1. Traffic volume changes due to general trends or to the countermeasure itself.
2. Changes in reported crash experience due to changes in crash reporting practice, weather, driver behavior, effects of safety programs, etc.
3. Regression-to-the-mean is problematic because safety is expected to change even in the absence of a treatment. A comparison group study will not account for regression-to-the-mean unless treatment and comparison sites are matched on the basis of crash occurrence.

In both the comparison group and empirical Bayes before-after methods, untreated sites are used to account for time trends and changes in other factors such as traffic volumes and crash reporting. As such, it is desirable to conduct a test of comparability to evaluate the suitability of the untreated group.

In literature different studies were carried out using the Empirical Bayesian before after analysis.

4.2.3. Full Bayes before-after analysis

Full Bayes is not a type of evaluation study on its own. Rather, it is a modeling approach that can be used in the same way as the more common generalized linear modeling approach, typically employed in the empirical Bayes method for before-after studies or in the development of cross-sectional models.

In the empirical Bayes approach, the prior information comes from using a reference group of sites similar to those under evaluation to calculate a sample mean and variance, or from a calibrated

safety performance function that relates the crash frequency of the reference sites to their characteristics. The point estimates of the expected mean and the variance are then combined with the site-specific crash count to obtain an improved estimate of a site's long-term expected crash frequency.

In the full Bayes approach, the prior information again comes from a model of a reference population but in this case, instead of a point estimate of the expected mean and its variance, a distribution of likely values is generated. This distribution of likely values is then combined with the site-specific accident frequency to obtain the estimate of long-term expected crash frequency. Through the use of a prior distribution instead of a point estimate the variance can be calculated more accurately.

Full Bayes models offer a number of potential advantages:

- The application of an integrated procedure to obtain outcomes.
- The small sample properties of FB models may allow the estimation of valid crash models with smaller sample sizes.

- The ability to include prior knowledge on the values of the coefficients in the modeling along with the data collected.
- The ability to consider spatial correlation between sites in the model formulation.
- The ability to specify very complex model forms.
- The ability to provide the posterior distributions of outcomes.

Regarding the last bulleted point, the FB method can accommodate distributions such as the hierarchical Poisson-Gamma distribution and the Poisson-LogNormal distribution, while the EB approach relies on the assumption of a negative binomial (NB) distribution of crash counts in using the NB dispersion parameter directly in the estimation process [4.12].

The principle issue with the full Bayes method is the complexity of its application as it may require a very high level of statistical training. Moreover, while it has been possible to develop software for application of the empirical Bayes method (e.g., Safety Analyst [4.14], this seems to be very difficult for the full Bayes method. Whether the benefits of the full Bayes method outweigh the increased complexity remains an open question.

Limited research to date suggests that the empirical Bayes approach will produce equally reliable results as the full Bayes method where sufficient sites are available to estimate robust safety performance functions for the empirical Bayes approach.

4.2.4. Cross sectional studies

Cross-sectional studies look at the crash experience of locations with and without some feature and then attribute the difference in safety to that feature. In its most basic application, the CMF is estimated as the ratio of the average crash frequency for sites with and without the feature. For this approach to be reliable it is important that all locations are similar to each other in all other factors affecting crash risk. In practice this requirement is difficult to meet.Cross-sectional studies are particularly useful for estimating CMFs where there are insufficient instances where the treatment was applied to conduct a before-after study. For example, there may be few or no projects where the shoulder is widened from, say, four feet to six feet. However, there would be many road segments with four foot shoulders and many with six foot shoulders. The reason that before-after studies are impractical in such cases is that there are often not enough before-after situations to allow for credible results.

In practice, it is difficult to collect data for enough locations that are alike in all factors affecting crash risk.

Hence, cross-sectional analyses are often accomplished through multiple variable regression models. In these models an attempt is made to account for all variables that affect safety. If such attempts are successful, the models can be used to estimate the change in crashes that results from a unit change in a specific variable. The CMF is derived from the model parameters.

The regression approach for estimating a CMF is consistent with the belief that the CMF is a function of the traits of the treated unit. A cross-sectional approach can be used to develop a CMFunction, and is preferable if the cause-effect relationship with crashes can be determined with confidence.

Bonneson et al. [4.2] developed a CMF using a set of cross sectional data to study the influence of curve radius in a rural two lane highways in Texas because in that case the execution of experiments and before-after studies were not practical or feasible although these latter give the best results. The CMF was developed as a function of the curvature radius, using a multivariate regression analysis, finding a different model form for the CMF developed by Harwood et al. [4.8] and reported in HSM.

$$CMF_{cr} = \frac{1{,}55L_c + \dfrac{80{,}2}{R_c} - 0{,}012I_S}{1{,}55L_c} \quad (4.15)$$

where,

- CMF_{cr} = horizontal curve radius accident modification factor;
- L_c = length of horizontal curve (= Ic Rc / 5280 / 57.3), mi;
- I_c = curve deflection angle, degrees;
- I_S = spiral transition curve presence (= 1.0 if spiral present, 0.0 if not present); and
- R_c = curve radius, ft.

Three different datasets were used in the elaboration, performing a Matched Pairs procedures for road segment. Hauer

[4.4] discusses the issues and challenges of using matched pairs in cross-section studies. A “Segment pair” dataset was carried out in which the segment pairs were identical, except that they had exhibit some variation in the input variables associated with the subject CMF. The segment-pair (SP) database was restructured to form two additional databases. The first database was used for regression model calibration. It was referred to as the “SP regression” database. In this database, each individual segment represents one observation. Thus, if there are n segment pairs, there would be 2n observations in the SP regression database. The second database is used for CMF calibration. It was referred to as the “SP group” database. In this database, the number of observations (or groups) is equal to the number of unique combinations of the “before” and “after” values for the geometric element of interest.

In this specific case the pairs are made up by the curves, that are investigated, and the adjacent tangent with an offset between the two element called “buffer zone” to avoid possible spatial correlation in the two element. This latter dataset contained the “After” condition, the tangent, and the “Before” condition, the curves.

All total, 3514 segments (1757 curved and 1757 tangent) were identified for the segment-pair database. They represent a total of 335.4 rural two-lane highway miles. Also, 1382 segments (691 curved and 691 tangent) were identified for the SP regression database as a subset of the segment-pair database. The smaller number of segments in the SP regression database is due to the use of the minimum exposure criterion for the selection of the element to be included in the analysis. These segments represent a total of 152.2 miles for three years of analysis from 1999 to 2001.

The segments in the SP regression database were associated with 566 crashes, of which 257 occurred on the tangent segments and 309 occurred on the curved segments. The segments in the segment-pair database were associated with 822 crashes, of which 349 occurred on the tangent segments and 473 occurred on the curved segments.

A regression analysis with an Empirical Bayesian correction was then applied on the SP regression dataset and used as “After” period. In this way the general model form for the CMF is the following:

$$X_{cr} = c_0 y E[N|X]_{\tan} CMF_{cr} \tag{4.16}$$

A second regression analysis on the segment pair dataset was used to estimate the regression parameter of the CMFunction on equation 3.16 considering the Harwood model form CMF for Curves. In this case a Nonlinear Regression procedure (NLIN) in the SAS software was used to estimate the calibration model coefficients:

$$CMF_{cr} = 1 + \left(d_0 + \sum_{i=1}^{5} d_i \operatorname{Region}_i \right) \cdot \frac{80,2}{1,55 L_c R_c} \tag{4.17}$$

where,

- d_i = calibration coefficients (i = 1, 2, 3,...);

- d_0= calibration coefficient corresponding to different region.

The goodness of fit of the regression models were validated using the Pearson χ^2 statistic, the root mean square error s_e useful for describing the precision of the model estimate, the coefficient of determination R^2 and the dispersion-parameter-based coefficient of determination R_k^2 developed by Miao [4.15].

The results of regression analysis are shown in the following equation:

$$CMF = 1 + 0,106 \cdot \left(\frac{5730}{R_c} \right)^2 \quad (4.18)$$

A sensitive analysis was then conducted on the CMF in comparison with the Harwood one. The revised CMF model is shown in Figure 3.2 for a range of radii. The values obtained from the revised model are shown with a solid trend line. The values obtained from the Harwood model form are shown using two dashed lines. This equation is sensitive to curve length; however, it was converted to include a sensitivity to curve deflection angle I_c instead by using the relationship between curve length and deflection angle provided in the variable definitions associated with the Harwood model form. This conversion was performed to facilitate a more equitable presentation of Harwood model form for the range of radii shown in Figure 3.2.

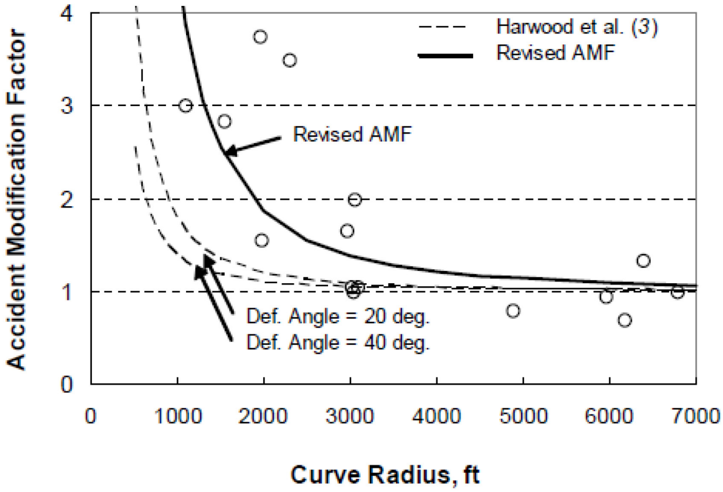


Figure 4.2. Relationship between Radius and CMF Value

The Authors concluded that, the results is consistent with another curve CMF previously derived using data from Washington. However, the application of cross sectional studies is challenging because it is not always possible to find a sufficient number of "identical" segments while maintaining the minimum sample size needed for statistical significance and that further research is needed to determine if there are any statistical implications (e.g., bias, loss of power) associated with the CMF estimate due to the use of the same data to calibrate the multivariate model and the CMF calibration model.

4.3. Application of the Before-After EB to estimate a CMF for safety barrier meeting a new EU Standard: a case study

A new EU regulation for safety barriers, which is based on performance, has encouraged agencies to perform an upgrade of the old barriers with the expectation that there will be safety benefits at the treated sites. The new class of barriers was designed and installed in compliance with the EN 1317 standards for Road Restraint Systems created in 1998 which lays down common requirements for the testing and certification of road restraint systems in all countries of the European Committee for Standardization, (CEN). Both the older and the new barriers are made of steel and are installed in a way to avoid vehicle intrusion, but the older ones are thought to be only effective at low speeds and low angles of impact. The new standard seeks to remedy this by providing better protection at higher speeds. The calibration of the CMF seeks to quantify the effect on the frequency of crashes (fatal+injuries) of retrofitting motorways with barriers meeting the new standards.

The new class of EU barriers are designed and installed in compliance with the European Norm (EN) 1317 standards. The EN 1317 for Road Restraint Systems was created in 1998 and lays down common requirements for the testing and certification of road restraint systems in all countries of the European Committee for Standardization (CEN), i.e. the 27 member states of the European

Union as well as Croatia, Iceland, Norway, Switzerland and Turkey. The EN1317 standard contains 5 parts relevant to roadside design guidelines:

- Part 1: Terminology and general criteria for test methods;
- Part 2: Performance classes, impact test acceptance criteria and test methods for safety barriers and vehicle parapets;
- Part 3: Performance classes, impact test acceptance criteria and test methods for crash cushions;
- Part 4: Performance classes, impact test acceptance criteria and test methods for terminals and transitions of safety barriers;
- Part 5: Product requirements and evaluation of conformity for vehicle restraint systems.

By comparison, the US standards, as outlined in NCHRP report 350 [4.16], have a similar approach to defining test levels by applying standard tests involving different vehicle types, impact speeds and impact angles. The impact speeds for heavy vehicles used in the European barrier tests are not as high as in the US tests, but this is compensated for by using heavier vehicles. Additionally, the European standard provides a vehicle occupant safety indication (impact severity levels) and working width (dynamic deflection) for each tested barrier. Impact severity levels relate to the degree of physical strain on the passengers depending on the values measured by Acceleration Severity Index (ASI) and Theoretical Head Impact Velocity (THIV). The evaluation factors taken into account by the US Standard are the following:

- Structural adequacy;
- Occupant risk;
- Vehicle trajectory.

For each of the evaluation factor there are a series of tests to be applied to establish the category of the barrier.

From 1998, EN1317 standards have been continuously reviewed and subjected to change. This study refers to road safety barriers placed in 2005 complying with the EN1317.2 in force from 2004, which is not substantially different from the present 2010 edition. The old barriers that were replaced in 2005 were not classified by any standard because they were placed during the A18 motorway construction in the years 1965 – 1971. Figure 4.3 provides examples of the old barrier on embankment (a) and on a bridge (b) and new barriers on embankment (c) and bridge (d).

The two barrier types can be compared based on the maximum containment level (CLmax):

$$CL = \frac{1}{2} M \cdot (V \cdot \sin \Theta)^2 \quad (4.19)$$

where:

- M: vehicle weight impact speed [kg];
- Θ: impact angle; and
- V: impact speed [m/s].

(a)

(b)

(c)

(d)

Figure 4.3. Pictures of the old barriers on embankment (a) and on a bridge (b) and the new ones on embankment (c) and on a bridge (d) installed on a motorway in Italy.

The containment level establishes the strength of the system, essentially specifying the maximum capacity for redirecting a vehicle. Higher containment levels produce stronger restraint systems. In Table 4.1 CLmax for old (Before) and new (After) barriers are reported highlighting the notable increase in the containment capacity of the new barriers placed in 2005.

Table 4.1. Values of the maximum containment level for old and new barrier for embankment and bridge sections

	CLmax (Before) [kJ]	*CLmax (After) [kJ]*	*Typology (After)**
Embankment	80	460	H3 (H4 median)
Bridge	150	570	H4a

** according to UNI EN 1317-2 2000 (EN 1317-2:1998)*

An investigation of the relationship between crashes and median barrier was carried out by Fitzpatrick et al. [4.17] [4.18] who developed a CMF from the coefficient of a regression model for Texas freeways that related crashes to the presence of a barrier and its offset from the edge of the carriageway. The results suggested a safety benefit for ran-off-road crashes, while, for the total number of crashes the impact on safety was negligible; for, small offsets, the results actually suggested an increase in the total number of crashes. Fitzpatrick, indeed, concluded that the CMF for rural freeway, based on data from Texas in which about 458 mi with barrier segment and 436 mi or rural area without barrier segment, give a real benefits in term of safety for the related crashes, while, for the total number of crashes this influence is negligible, or, for very small value of offset, give an increment of the total number of crashes. Elvik et al. in the “Handbook of road safety measures” summarized the results of different studies on the effect on running-off-the-road accidents of setting up guardrails along the roadside [4.19]. This meta analysis indicated strong reductions of 44% and 47% for fatal and injury crashes, respectively (Table 4.2).

Based on several studies, the CMF proposed in the HSM [4.1] suggests a reduction in the number of injury run-off-road accidents for changing the type of roadside barrier along an embankment to a less rigid type. Conversely, more rigid barriers (e.g. concrete or steel versus wire or cable) produce an increase in ran-off-road crashes of up to 40%, according to the HSM.

Table 4.2. Meta analysis on guardrail retrofitting from the "Handbook of Road Safety measures"

Accident Severity	Percentage change in the number of accidents		
	Types of accident affected	**Best estimate**	**95% confidence interval**
	New guardrail along embankment		
Fatal accidents	Running-off-the-road	-44	(-54, -32)
Injury accidents	Running-off-the-road	-47	(-52, -41)
Unspecified	Running-off-the-road	-7	(-35, +33)
	Changing to softer guardrail		
Fatal accidents	Running-off-the-road	-41	(-66, +2)
Injury accidents	Running-off-the-road	-32	(-42, -20)

A study by Scully et al. [4.20] in Australia indicated a 42.2% reduction in all casualty crashes, but, based on a literature review, Austroads (Table 4.3) [4.21] suggests that the installation of safety barriers results in an average reduction of 40% but only for run-off-road crashes.

Zegeer et al. [4.22] studied the effect of the distance of safety barriers from the edge of the travelled way (defined as clear zone) for two lane undivided rural roads. The results show reductions in ran off road crashes ranging from 13% for 1.5 m of clear zone to 44% for 6 m of clear zone.

In the Crash Modification Clearinghouse managed by FHWA [4.23], which contains over 3,000 CMF estimates for a wide range of safety countermeasures under a variety of conditions, 25 CMFs for “Countermeasure: Improve guardrail” are reported, with an average value of 0.82 (min=0.50; max=0.95) for all crash types and 0.75 for run off road and fixed object crash types (min=0.68; max=0.82).

The reference for these CMFs is a study of Gan et al. published in 2005 [4.24].

Table 4.3. Summary of studies on margin protection from AustRoads

Study	year	Country	Environment	Reduction
Zegeer et al.	1987	USA	2-lane undivided rural road	Reduction in run-off-road crashes with how many meters of clear zone 13% at 1.5 21% at 2.4 25% at 3.3 29% at 3.6 35% at 4.5 44% at 6
Beca Ltd	1998	NZ	Median barriers	75% reduction
Elvik and Vaa	2004	Netherlands	Gurdrail along embankment	-7% run-off-road crashes
			Gurdrail on median	+24% run-off-road crashes
Scully et al.	2006	Australia	Gurdrail installation	42.2% casualty crashes

In summary, despite differences in the absolute value of CMF, all the above studies indicate that road guardrails are effective in reducing target crashes, but they can have also negative effects on other types of crashes.

The safety effectiveness of the barrier is also related to the barrier type and positioning. However, these CMFs may not be appropriate for the scenario under consideration in this study. For example, the only ones in the CMF clearinghouse related to an improvement of guardrail, are not specific to motorways, and in particular not to motorway barriers meeting the new EU standard.

4.3.1. Segmentation approach and data treatment

For the elaboration an empirical Bayes Before/After analysis was chosen to address the problem related to the regression to the mean effects. The data described in the chapter one, as well as the segmentation approach used for the CMF calibration are reported below.

The segmentation approach chosen in this study, was carried out in a way to use the barrier typology as the only homogenous variable. In other terms all the segment are homogeneous respect to barriers typology, in this way the presence of old or new barriers is implicit considered in the models.

The advantage of using homogeneous segment respect to a dummy variable in the regression model is that they do not impose any preconceived functional relationship on the estimates obtained from the model. If a functional relationship is used in the regression model, then this relationship could be reflected in the expected values used to calibrate the CMF. It follows that, if a functional relationship is used in the model, then it could indirectly bias the CMF calibration. The use of homogeneous segment respect of the variable investigated minimizes the potential for this type of bias because it does not require the specification of a function for the subject CMF input variable.

The other variable consider in the models, the curvature, was treated to adapt it to the segmentation approach chosen. This happened because it is a continuous variables.

In particular it was treated as curvature change rate (CCR) [4.25] of the segment, calculated as follows:

$$CCR = \frac{\sum_i |\gamma_i|}{L} [gon/m] \qquad (4.20)$$

where γi is the deflection angle for a contiguous element (curve or tangent) i within a section of length L;

The segments are homogeneous, excluding the barrier typology, respect to the cross section typology, in this case viaduct Embankment and Trench. So each segment was characterized with the typology of cross section. This was made possible because in the segmentation approach based on barrier typology because each segment was characterized by a unique value of cross section typology.

The data used for this investigation are based on an Italian rural motorway, the “A18” Messina-Catania, which is approximately 76 km (47.2 miles) long. The cross section is made up of four 3.75 m lanes, 2 in each direction, 3.75 m plus an emergency lane that is 3.00 m wide. Carriageways are divided by a median with barriers.

The analysis period is for the 12 years from 2000 until 2004 for the before period and form the 2006 through 2012 for the after period within which 418 severe (fatal plus injury) crashes according to the official statistics on motor vehicle collisions provided by the Italian National Institute of Statistics (ISTAT) [4.25]. For the elaboration all the data were divided into three different database. The first database represent the group of segment used as reference, in which the treatment was not present both in the before and after period, called reference group dataset, Table 3.4 reports the traffic and crashes statistic over the years of analysis for the total crashes.

In Figure 4.4 and 4.5 the traffic distribution is reported, for the before period (a) and the after period of analysis (b).

The others groups were made up by the segment in which the barriers were changed in the 2005, and are related to the before and the after period. The dataset related to the before and after period of the treatment contain segment which the median, the lateral or both were changed. Table 4.5 and 4.6 report the basic information about the before and after dataset used in the elaboration.

Table 4.4. AADT and crash data related to the total number of crashes (RG=Reference Group, TG=Treated Group)

	Year	Range AADT	Injury Crashes		Injury Crashes/(Million veh*km)		Total Length (km)	
			TG	RG	TG	RG	TG	RG
Total	2000	10577 – 32998	16	40	0.29	0.1	16	58
	2001	10662 – 35799	17	35	0,36	0.12		
	2002	8696 – 24904	7	37	0.31	0.11		
	2003	9082 -26123	10	49	0.23	0.11		
	2004	9423 – 26947	11	46	0.24	0.09		
	2006	7792 – 26414	5	43	0.03	0.08		
	2007	7917 – 27001	12	37	0.19	0.10		
	2008	7651 – 26783	6	30	0.13	0.10		
	2009	9066 – 26743	3	31	0.09	0.2		
	2010	10622 – 37052	13	26	0.20	0.13		
	2011	10262 – 36375	9	24	0.16	0.12		
	2012	9294 – 34174	12	20	0.18	0.13		
		Total	**121**	**418**	**0.17**	**0.11**		

It is evident that there was a treatment site selection bias, in that sites with higher crash rates tended to be selected for treatment. This would result in regression to the mean and necessitates the use of the empirical Bayes methodology used in this study to account for this bias.

The analysis were conducted on the total number of crashes, the direct related category of crashes, ran-off-road crashes, and the non-ran-off-road crashes. The ran-off-road crashes are the direct related crashes to the roadside condition, and in this study to the barrier typology. Table 4.5 and 4.6 reports the traffic and crash data distribution for these category of crashes

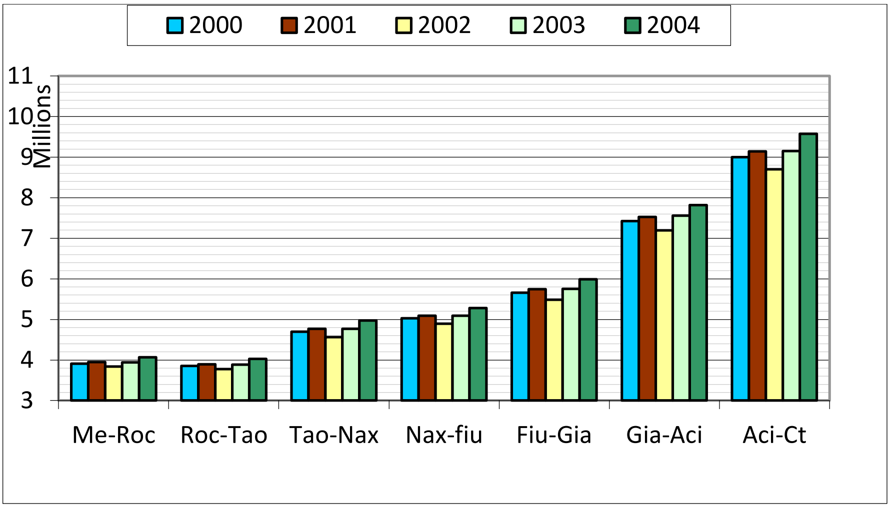


Figure 4.4. traffic distribution for the before period

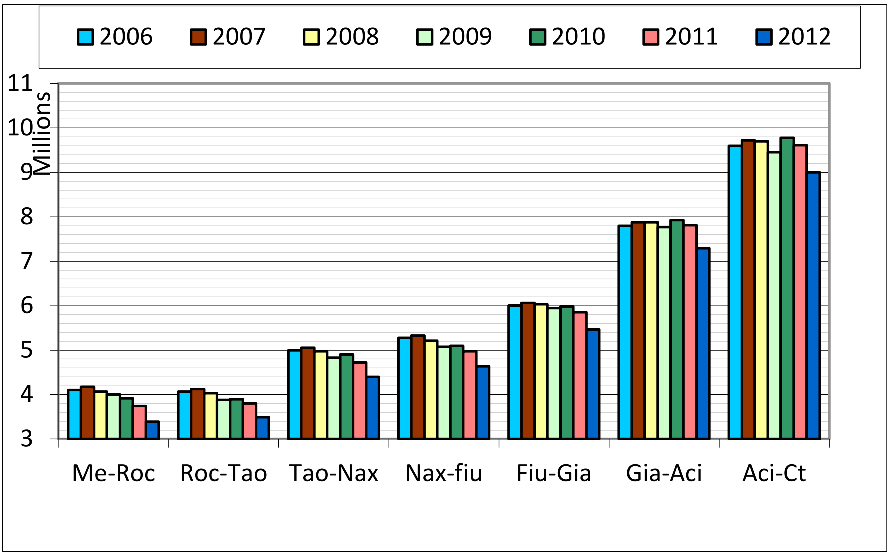


Figure 4.5. traffic distribution for the after period.

In this study, only the road segments were analyzed; interchange data and the part of segment directly influenced by the presence of intersection were discarded. Every segment contiguous to an intersection starts from a distance of 50 m (164 ft) from the bevel for the insertion of the service lanes for exit from, and entry into the main flow. The data available, in addition to AADT , were: radius of curvature, vertical gradient, type of section, and roadside features (presence and typology of the lateral and median barriers), although only curvature (CCR) was used in the calibration of CMF.

Table 4.5. AADT and crash data related to the ran-off-road crashes (RG=Reference Group, TG=Treated Group)

	Year	Range AADT	Injury Crashes		Injury Crashes/(Million veh*km)		Total Length (km)	
			TG	RG	TG	RG	TG	RG
Ran off road	2000	10577 – 32998	14	20	0.27	0.06	16	57
	2001	10662 – 35799	10	13	0.20	0.04		
	2002	8696 – 24904	5	17	0.3	0.05		
	2003	9082 -26123	3	24	0.05	0.03		
	2004	9423 – 26947	8	21	0.18	0.05		
	2006	7792 – 26414	2	23	0.01	0.04		
	2007	7917 – 27001	8	21	0.01	0.07		
	2008	7651 – 26783	2	15	0.04	0.07		
	2009	9066 – 26743	2	16	0.03	0.13		
	2010	10622 – 37052	4	20	0.05	0.15		
	2011	10262 – 36375	2	13	0.03	0.12		
	2012	9294 – 34174	2	11	0.02	0.11		
		Total	**62**	**214**	**0.1**	**0.06**		

The traffic distribution in the two travel directions were analyzed to avoid problem related to consider large difference in the exposure factor among the segments.

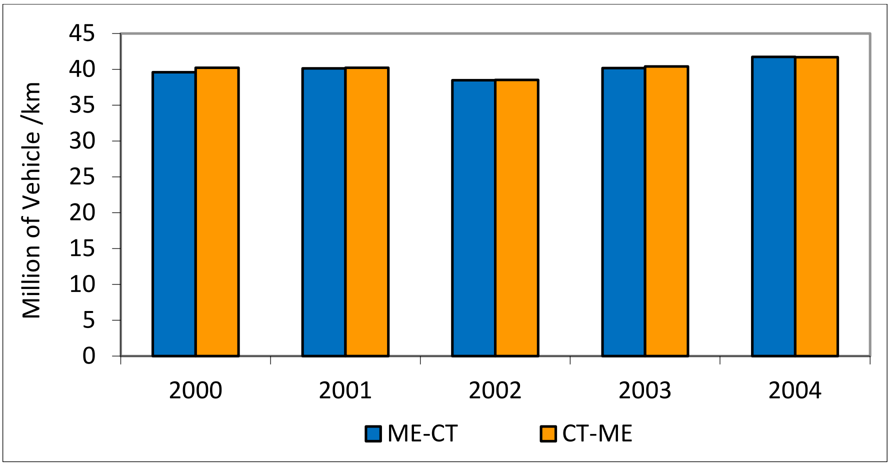
Figure 4.6 shows the traffic distribution in the period of analysis for the before period (a) and the after period (b). There are not significant differences between the traffic distribution in the two directions.

As earlier in the Chapter, the EB analysis requires the use of a safety performance function (SPF).

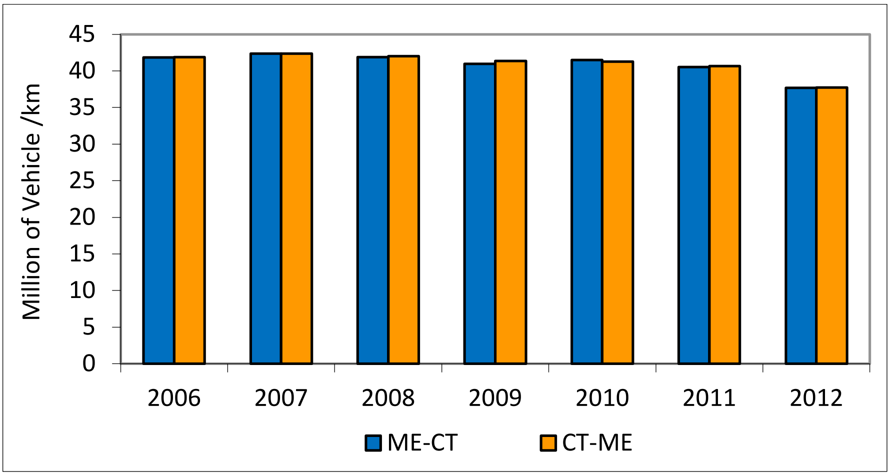
The Generalized Linear Modeling (GLM) [4.28] [4.29] approach was used to estimate the SPFs, using the Statistical Analysis System (SAS) [4.30] software package.

Table 4.6. AADT and crash data related to the non-ran-off-road crashes (RG=Reference Group, TG=Treated Group)

	Year	Range AADT	Injury Crashes		Injury Crashes/(Million veh*km)		Total Length (km)	
			TG	RG	TG	RG	TG	RG
Non-Ran off road	2000	10577 – 32998	2	20	0.01	0.05	**16**	**57**
	2001	10662 – 35799	7	22	0.02	0.06		
	2002	8696 – 24904	2	20	0.01	0.06		
	2003	9082 -26123	7	25	0.18	0.08		
	2004	9423 – 26947	3	25	0.06	0.04		
	2006	7792 – 26414	3	20	0.02	0.04		
	2007	7917 – 27001	4	16	0.18	0.03		
	2008	7651 – 26783	4	15	0.09	0.03		
	2009	9066 – 26743	1	15	0.06	0.07		
	2010	10622 – 37052	9	16	0.05	0.03		
	2011	10262 – 36375	7	24	0.04	0.05		
	2012	9294 – 34174	10	9	0.06	0.06		
		Total	**59**	**227**	**0.09**	**0.05**		



(a)



(b)

Figure 4.6. Traffic distribution among the two travel direction for the before period *(a)* and for the after period *(b).*

4.3.2. Analysis and results

The first step in the analysis was to validate the reference group using the methodology developed by Hauer [4.3], and described above for the suitability of the dataset. The results are shown in Tables 4.7 to 4.9 for the total number of crashes, the ran-off-road and the non-ran-off-road respectively.

The conditions requested by the test were that mean has to be close to one and the 95% confidence interval have to include 1.

The analysis of the results of the test on the three categories of crashes lead to the conclusion that the reference group is suitable for the total crashes and for the ran-off-road crashes.

Different considerations have to be done for the non-ran-off road crashes.

The 95% interval confidence doesn’t include 1 and the mean is not so close to 1.

On the other hand the upper value of the interval is close to 1 and the test is only suggested for the empirical Bayes analysis, while is fundamental for the before-after analysis with comparison group described earlier in the Chapter.

As such, the results of the test have to be seen as a validation of the Reference Group in term of regression to the mean effect and selection bias. To consider time trend in the calibration of CMF, a post SPF calibration procedure was applied. The SPF was calibrated considering an average AADT value for the whole period of analysis (12 years) and the sums of crashes for each segment. The estimation obtained for each year was than corrected with a multiplier given by the ratio of the sums of yearly observed crashes and the SPF estimates for the reference sites.

Table 4.7. Results obtained from the test of comparability on the reference Group for the total crashes.

Total			
Year	**Treatment Group**	**Reference Group**	**o**
2000	16	40	-
2001	17	35	0.75
2002	7	37	2.19
2003	10	49	0.82
2004	11	46	0.76
2006	5	43	1.68
2007	12	37	0.32
2008	6	30	1.35
2009	3	31	1.51
2010	13	26	0.17
2011	9	24	1.15
2012	12	20	0.55
	Mean		**0.93**
	s^2		0.34
	Var(ω) tot		0.08
	Var(ω)		0.21
	95% confidence interval		
	1.09	**0.76**	

Consistent with the state of research in developing these models, the negative binomial error distribution was assumed for the count of observed crashes [4.4]. For the empirical Bayes evaluation, the negative binomial dispersion parameter was estimated from the calibration of the SPF using a maximum likelihood methodology.

Table 4.8. Results obtained from the test of comparability on the reference Group for the ran-off-road crashes.

Ran-off-road			
Year	**Treatment Group**	**Reference Group**	**o**
2000	14	20	-
2001	10	13	0.79
2002	5	17	2.04
2003	3	24	1.69
2004	8	21	0.28
2006	2	23	2.83
2007	8	21	0.19
2008	2	15	1.8
2009	2	16	0.68
2010	4	20	0.47
2011	2	13	0.83
2012	2	11	0.53
Mean			**1.04**
s^2			0.66
Var(ω) tot			0.11
Var(ω)			0.41
95% confidence interval			
1.26		**0.82**	

Some recent studies [4.31] [4.32] [4.33], suggested that the dispersion parameter, contrary to earlier research, is not constant for a given data set but actually varies from site to site depending on the length of a roadway segment. The varying form in applications such as the Highway Safety Manual [4.1] is such that the dispersion parameter for certain classes of road segments is inversely proportional to segment length.

Table 4.9. Results obtained from the test of comparability on the reference Group for the non- ran-off-road crashes.

non-Ran-off-road			
Year	**Treatment Group**	**Reference Group**	**o**
2000	2	20	-
2001	7	22	0.26
2002	2	20	2.05
2003	7	25	0.29
2004	3	25	1.69
2006	3	20	0.58
2007	4	16	0.46
2008	4	15	0.71
2009	1	15	1.93
2010	9	16	0.10
2011	7	24	1.60
2012	10	9	0.22
Mean			**0.85**
s^2			0.57
Var(ω) tot			0.04
Var(ω)			0.31
95% confidence interval			
0.99		**0.76**	

This form was first suggested by Hauer [4.34], who argued logically that shorter segments have a higher accident frequency variance and, consequently, should have a higher dispersion parameter than longer segments, and that this variation should influence the long-term estimate of a segment's safety. For this study, following Cafiso et al. [4.35], proportionality was not, a priori, assumed. Instead, the chosen equation for the calibration of the variable dispersion parameter was the following (Equation 4.21)

$$k = \alpha \cdot L^{\gamma} \tag{4.21}$$

where α and γ are regression terms.

Models based on the dispersion parameter varying with length were calibrated with a modified negative binomial regression technique, in which the dispersion parameter in the log-likelihood function was considered to vary according to the previously established exponential form (Equation 4.21).

The maximization of the log-likelihood function required an iterative process to calibrate both the SPF coefficients and the two coefficients for the exponential function for k [4.35]. To this end, an iterative calculation algorithm was developed and implemented. The results of the model calibration for the three crash types are shown in Table 4.10, based on the SPF model form shown in Equation 4.22, while Figures 4.6 to 4.8 report a measure of the goodness of fit of the models.

$$E(Y) = y_i \times e^{\alpha} \times L \times AADT^{\beta} x e^{\beta_1 x CCR} \tag{4.22}$$

where:

- E (Y): expected annual (fatal plus injury) crash frequency of random variable Y;
- L: length of road segment [m];
- AADT: average annual daily traffic [veh/day];
- α, β and β_1 are regression terms;
- yi is the time trend coefficient in the year i

To evaluate the goodness of fit (g.o.f.) of the models, the Cumulative Residuals (CURE) plot [4.36] method was applied.

The CURE method is based on the study of residuals, i.e., the difference between the number of crashes observed at a site and the expected value at the same site and in the same year.

Table 4.10. Value of regression parameters, (p-value) and [Standard error] for the SPFs calibrated.

		Total	Ran off road	Non -Ran off road
Intercept		-14.1368 (2.153) [<.0001]	-13.3460 (2.444) [<.0001]	-16.377 (2.5830) [<.0001]
AADT		0.9631 (0.224) [<.0001]	0.7862 (0.254) [0.002]	1.1416 (0.268) [<.0001]
CCR		2.281 (0.004) [<.0001]	2.486 (0.0236) [<.0001]	2.623 (0.006) [<.0001]
Years	2000	1.05	1.33	0.88
	2001	0.91	0.85	0.95
	2002	1.00	1.15	0.90
	2003	1.27	1.58	1.07
	2004	1.16	1.35	1.04
	2006	1.08	1.47	0.82
	2007	0.82	1.07	0.65
	2008	0.75	0.95	0.61
	2009	0.79	1.03	0.62
	2010	0.91	1.29	0.67
	2011	0.72	0.85	1.00
	2012	0.55	0.76	0.42
k		$5.6x\boldsymbol{L}^{-0,8}$	$6.1x\boldsymbol{L}^{-0,85}$	$5.5x\boldsymbol{L}^{-0,375}$

Assuming that residuals are normally distributed with expected value equal to 0 it is possible to calculate the variance σ of the expected value as the square of the cumulate residuals [4.36]. Table 4.11 reports the Crash Modification Factors calibrated on the data for the total, ran-off-road and non-ran off road crashes.

The plot of cumulative residuals should oscillate around 0, and not exceed the ±2σ* bounds.

As is evident from Figures from 4.6 through 4.8, the cumulative residuals plots show a reasonable good fits of the model to the datasets. The estimates indicate a strong enough safety benefit for ran-off road crashes, without any change in non-ran-off road crashes.

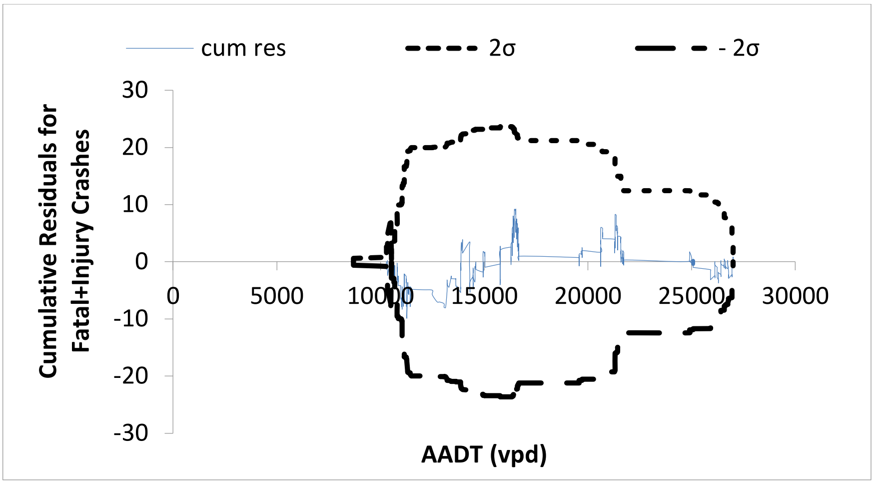


Figure 4.6 CURE Plots with ± 2σ for the total crashes.

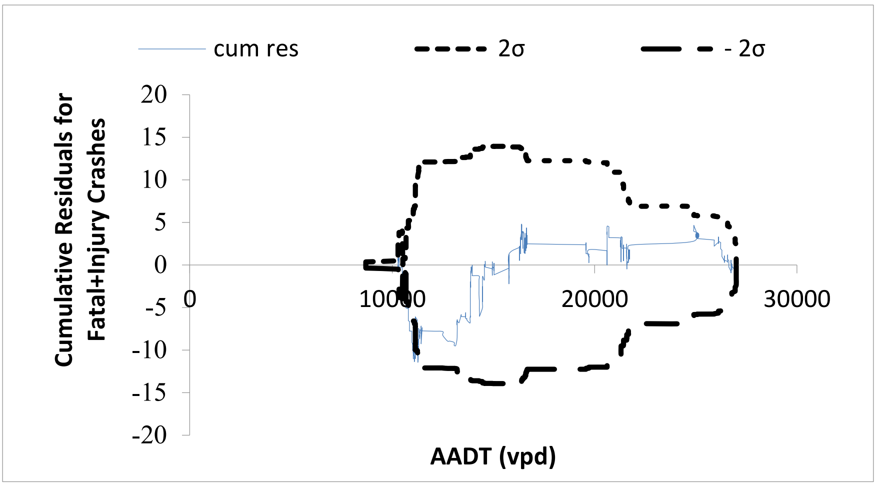


Figure 4.7 CURE Plots with ± 2σ for the ran-off-road crashes.

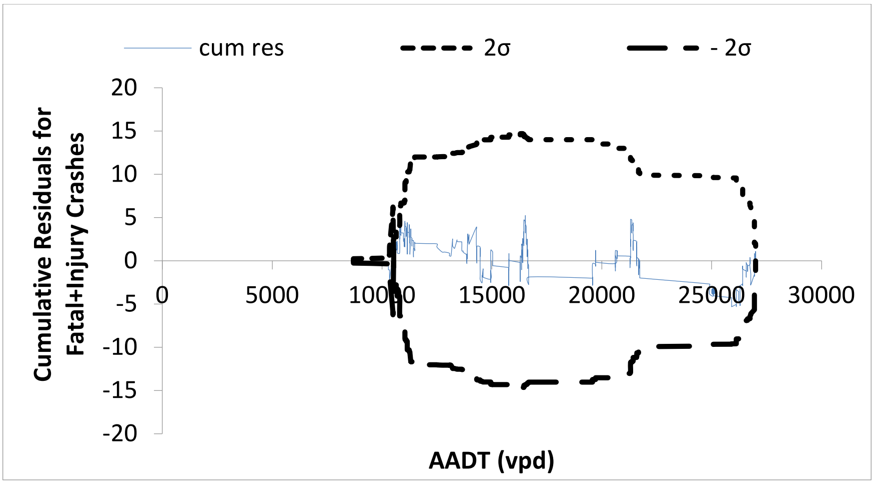


Figure 4.8 CURE Plots with ± 2σ for the non-ran-off-road crashes.

Table 4.11. CMFs estimation results for fatal plus injury crashes.

	Total		Ran off road		Non-Ran off road	
Comp. Ratio (*av. value*)	0.93		0.87		1.05	
Var (Bsum)	24.85		31.95		5.68	
Bsum	96.78		76.40		47.52	
Asum	69		22		47	
CMF	***0.71***		***0.28***		***0.98***	
Stdev	0.09		0.065		0.152	
95% interval	***0.52***	***0.79***	***0.16***	***0.41***	***0.68***	***1.28***

The reduction found in the total crashes is due to the higher percentage of ran-off-road crashes in comparison with the other crash categories. If the whole period of analysis is taken into account the ran-off-road crashes in the reference group are about the 52% of the total crashes.

The percentage in the before period is about the 45% and 55% in the after period for the reference group. In the treated sites the percentage are 65% and 35% in the before and after period respectively. The standard deviation associated to the CMF indicates a strong benefit for the ran-off-road crashes and for the total crashes with both the upper limit of the 95% confidence interval smaller than one. The CMF calibrated on the ran-off-road crashes give the best results in term of standard deviation too, with the lower value, that indicates a good size of the sample. On the contrary the large standard deviation of the CMF calibrated on the non-ran-off-road crashes indicates that the sample of treated sites is not yet large enough to estimate a robust CMF with sufficient statistical significance for that categories of crashes.

4.4. Crash Modification Function

If the cross section distribution of the CMF is considered, the effects of the CMF could be related to one or more road feature present in the segments, and in particularly to those which from an engineering point of view should have a functional relationship with the treatment investigated.

In this case, speaking about lateral protection, and taking into account the available variable the curvature change rate was considered.

Calculating the value of the CMF in the treated sites, according to the segmentation chosen, considering for each of them the expected value and the variance it was possible to make a regression analysis on the dataset finding a relationship between the CMFs and the curvature change rate value.

The curvature change rate give an idea of the deflection of the segment on its length. Shorter segment with the same curves (length and radius) of longer one have a higher value of CCR, so its value is strongly influenced by the segmentation approach. In the specific case of a segmentation based on barrier typology the CCR results independent from others variable and it is the only way to avoid short segment due to a segmentation approach based both on barriers and curvature.

The problem of short segment is not only related to the goodness of fit of the model to the dataset, but also to the calibration of the cross site CMFs. In the mathematical expression indeed, the observed crashes in the after period are used to make a comparison with the expected crashes in the after period, in the same sites, if the treatment had not been applied. From an operative point of view it is impossible to calculate the CMF in a site if no crashes have occurred in the after analysis period or better if data have a high reliability and there is not regression to the mean bias the CMF produces a 100% of reduction of crashes without variance. By the way the random nature of crashes and the effect of the RTM bias can easily disprove this statement.

To overcome that problem a cluster analysis is performed. In the following the description of the methodology used in the elaboration and the results.

4.4.1. Cluster analysis

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

Besides the term clustering, there are a number of terms with similar meanings, including automatic classification, numerical taxonomy, botryology (from Greek βότρυς "grape") and typological analysis. The subtle differences are often in the usage of the results: while in data mining, the resulting groups are the matter of interest, in automatic classification primarily their discriminative power is of interest. This often leads to misunderstandings between researchers coming from the fields of data mining and machine learning, since they use the same terms and often the same algorithms, but have different goals.

Cluster analysis was originated in anthropology by Driver and Kroeber in 1932 and introduced to psychology by Zubin in 1938 and Tryon in 1939 [4.38] and famously used by Cattell beginning in 1943 [4.39] for trait theory classification in personality psychology.

According to Vladimir Estivill-Castro, the notion of a "cluster" cannot be precisely defined, which is one of the reasons why there are so many clustering algorithms [4.40]. There is a common denominator: a group of data objects. However, different researchers employ different cluster models, and for each of these cluster models again different algorithms can be given. The notion of a cluster, as found by different algorithms, varies significantly in its properties. Understanding these "cluster models" is key to understanding the differences between the various algorithms [4.41]. Typical cluster models include:

- Connectivity models: for example hierarchical clustering builds models based on distance connectivity.
- Centroid models: for example the k-means algorithm represents each cluster by a single mean vector.
- Distribution models: clusters are modeled using statistical distributions, such as multivariate normal distributions used by the Expectation-maximization algorithm.
- Density models: for example DBSCAN and OPTICS defines clusters as connected dense regions in the data space.
- Subspace models: in Biclustering (also known as Co-clustering or two-mode-clustering), clusters are modeled with both cluster members and relevant attributes.
- Group models: some algorithms do not provide a refined model for their results and just provide the grouping information.
- Graph-based models: a clique, i.e., a subset of nodes in a graph such that every two nodes in the subset are connected by an edge can be considered as a prototypical form of cluster. Relaxations of the complete connectivity requirement (a fraction of the edges can be missing) are known as quasi-cliques.

A "clustering" is essentially a set of such clusters, usually containing all objects in the data set. Additionally, it may specify the relationship of the clusters to each other, for example a hierarchy of clusters embedded in each other. Clusterings can be roughly distinguished as:

- hard clustering: each object belongs to a cluster or not
- soft clustering (also: fuzzy clustering): each object belongs to each cluster to a certain degree (e.g. a likelihood of belonging to the cluster)

There are also finer distinctions possible, for example:

- strict partitioning clustering: here each object belongs to exactly one cluster
- strict partitioning clustering with outliers: objects can also belong to no cluster, and are considered outliers.
- overlapping clustering (also: alternative clustering, multi-view clustering): while usually a hard clustering, objects may belong to more than one cluster.
- hierarchical clustering: objects that belong to a child cluster also belong to the parent cluster
- subspace clustering: while an overlapping clustering, within a uniquely defined subspace, clusters are not expected to overlap.

In the elaboration the algorithm used for the clustering is the median. The distance between the elements in the cluster and the median is calculated as an Euclid distance.

Table 4.12. Clustering of the dataset respect to CCR for the total and ran-off-road crashes.

Total			
Cluster Median	**Length (m)**	**Observed**	**Cluster**
1.1082	2823.495	4	1
27.3214	7798.79	23	2
58.7093	2991.85	28	3
80.6199	3079.73	8	4
125.456	484.41	0	5
Ran-off-Road			
Cluster Median	**Length (m)**	**Observed**	**Cluster**
1.1082	2823.495	1	1
27.3214	7798.79	10	2
58.7093	2991.85	4	3
80.6199	3079.73	2	4
125.456	484.41	0	5

In the analysis 68 segments are used for the clustering, each of them is pair with his own value of CCR. The clustering analysis is performed on the value of CCR. The results are shown in Table 4.12 for the total crashes and the ran-off-road crashes. The elaboration was performed using the Statgraphics software package [4.42].

As it is clear from table 4.12 the cluster number 5 is relatively small in term of length. To avoid problem due to the quality of data (i.e. localization of crashes) or RTM effects on the observed data the 4th and 5th cluster were merged. The new centroid was assumed equal to the median of the sample using the same criteria of the software was used to do this. In the table 4.13 are shown the results. In the next Paragraph a regression analysis on the cluster is performed to the aim to calibrate a function.

Table 4.13. Clustering of the data respect CCR merging the 4th and 5th clusters for the total and ran-off-road crashes.

Total			
Cluster Median	**Length (m)**	**Observed**	**Cluster**
1.1082	2823.495	4	1
27.3214	7798.79	23	2
58.7093	2991.85	28	3
103.038	3564.14	8	4 - 5
Ran-off-Road			
Cluster Median	**Length (m)**	**Observed**	**Cluster**
1.1082	2823.495	1	1
27.3214	7798.79	10	2
58.7093	2991.85	4	3
103.038	3564.14	2	4 - 5

4.4.2. Crash Modification Function calibration

Using the results of the clustering a regression analysis was performed on the CMFs calibrated for each cluster.

The advantage of having a function lies in the fact that it is not always reasonable to assume a uniform safety effect for all sites with different characteristics [4.3]. A countermeasure may also have several levels or potential values, for this reason developing Crash Modification Functions give the best fit to reality when an analysis on the effect of a treatment has to be investigated. A crash modification function allows the CMF to change over the range of a variable or combination of variables.

If the protection of margin is investigated the direct related typology of crashes are the ran-off-road. In general that crash typology is expressed as a loose of control of the vehicle that try to abandon the carriageway. As such the curvature plays a fundamental role in the causation of the crash events.

In Table 4.15 the length of segment is reported together with the cross site calibration of the CMF, the observed number of crashes and the curvature change rate value and the variance of the CMF for the total crashes for the total and ran-off-road crashes.

The CCR values reported on Table 4.15 is related to the median of the class used in the cluster analysis.

Using the CMFs, calibrated on each site, and CCR values a regression analysis was performed considering the variation of the CMF together with the variation of the CCR. The results are shown in the Figures 4.9 for a exponential regression on the total crashes and in Figure 4.10 and 4.11 for the linear regression and exponential regression, respectively, for the ran-off-road crashes.

The coefficient of determination (R^2) was used to assess the goodness of fit of the regression. The coefficient of determination, denoted R^2, indicates how well data points fit a line or curve. R^2 is a statistic that will provide some information about the goodness of fit of a model.

In regression, the R^2 coefficient of determination is a statistical measure of how well the regression line approximates the real data points. The R^2 value is a popular measure used to judge the adequacy of a regression model. Defined as a ratio, the R^2 value is a proportion that represents the variability of the dependent variable that is explained by the model. In symbolic form:

$$R^2 = \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (4.23)$$

where:

$\hat{y}$: model estimates,
$\bar{y}$: mean of the observations
y_i: actual observations

An R^2 value near zero indicates that there is no linear relationship between the dependent and independent variables, while a value near 1 indicates a linear fit.

Table 4.15. CMFs calibration results on cluster analysis for the total and ran-off-road crashes.

Total				
Cluster (CCR)	Length (m)	Observed	CMF	Cluster
1.1082	2823.495	10	0.65	1
27.3214	7798.79	23	0.49	2
58.7093	2991.85	28	1.28	3
103.038	3079.73	8	0.56	4 - 5
Ran-off-Road				
Cluster (CCR)	Length (m)	Observed	CMF	Cluster
1.1082	2823.495	6	0.33	1
27.3214	7798.79	10	0.32	2
58.7093	2991.85	4	0.26	3
103.038	3079.73	2	0.14	4 - 5

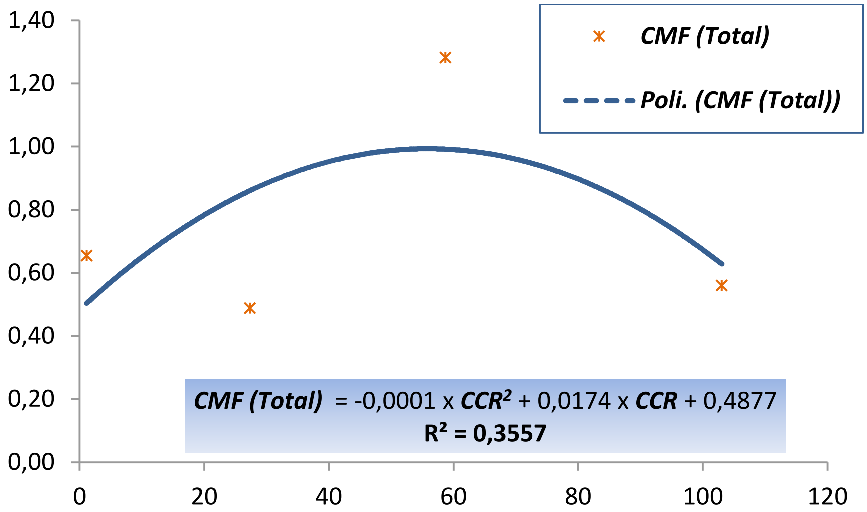


Figure 4.9. Exponential regression on the cross site distribution of the CMF on total crashes.

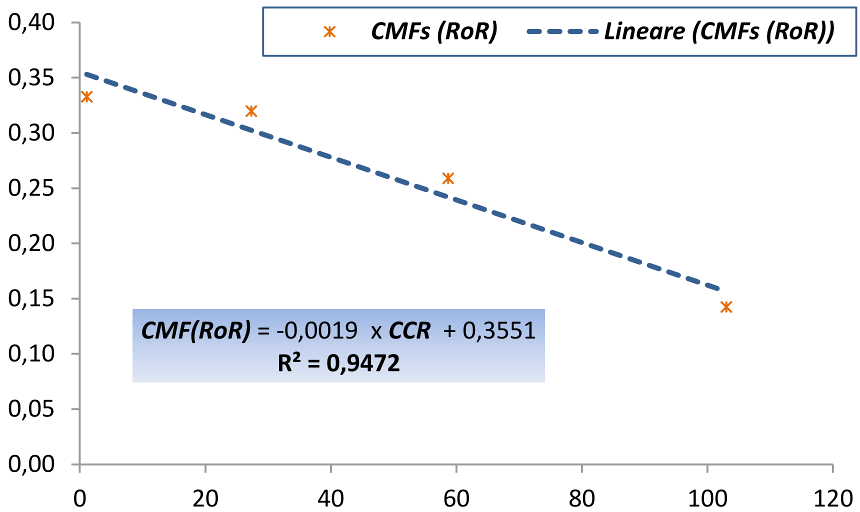


Figure 4.10. Linear regression on the cross site distribution of the CMF on ran-off-road crashes.

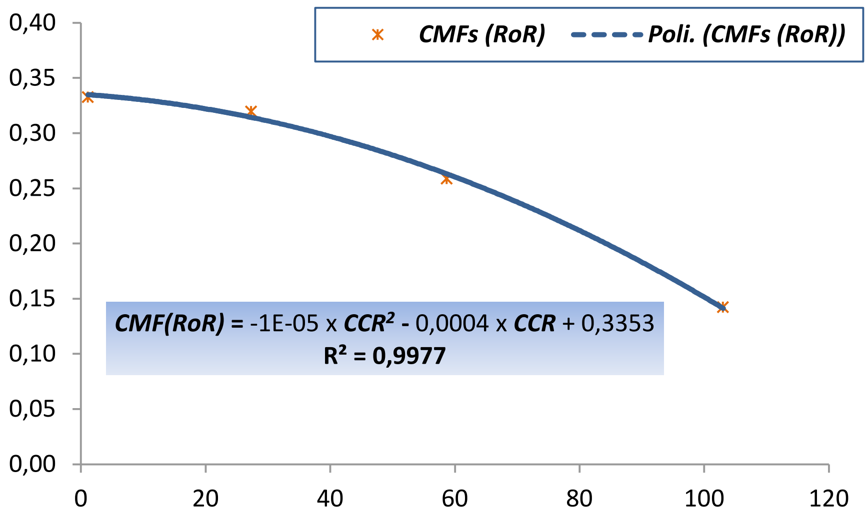


Figure 4.11. Exponential regression on the cross site distribution of the CMF on ran-off-road crashes.

From the analysis of the regression performance shown in the Figures 4.9 through 4.10 both the exponential regression and the linear one have a good goodness of fit to the cross site CMFs for the ran-off-road crashes. The exponential regression has the following expression:

$$CMF = -10^{-4} \cdot CCR^2 - 0.0004 \cdot CCR + 0.3353 \quad (4.24)$$

While the linear regression has the equation:

$$CMF = -0.0019 \cdot CCR + 0.3551 \quad (4.25)$$

From the analysis of the R^2 value, the exponential relationship gives the best results with a closer value to one.

Different consideration have to be made on the total crashes. The value of R^2, indeed, indicates that the performance of the regression is not enough to say that a relationship exist between CCR and the reduction of crashes due to the retrofitting motorways with barriers meeting the EU standard. It is due to the influence of the Non ran-off-road crashes on the total crashes. The ran-off-road are mainly single vehicle crashes, and the influence of the curvature is determinant in the causation of the crash event.

Generally the Crash Modification Functions are not provided with a variance and a standard deviation. In Bonneson et al. [4.2], in the HSM [4.1] and in the web based database for CMF, or rather in the Crash Modification Clearinghouse managed by FHWA [4.42], which contains over 3,000 CMF estimates for a wide range of safety countermeasures under a variety of conditions, the variance is not taken into account when CMFunction are reported or calibrated.

By the way for the analysis reported in the next Chapter 4 the variance of the Crash Modification Factor play a fundamental role in the Benefit/Cost Analysis, even if a Function, or a cross site distribution of the CMFs, is considered.

For this reason the value of the variance needs to be estimated for the cross site distribution of the CMF.

The methodology used to estimate the variance of the CMFunction has been proposed by Hauer et al. [4.37] who used a weighted variance of the combination of two CMFs related to the same treatment. In Chapter 5 a methodology for calculating a cross site variance is applied [4.37] together with a new methodology for the Benefit/Cost analysis.

4.5. Chapter summary

In the present Chapter different methodologies to estimate a CMF were introduced with a wide literature review. At the end of the Chapter study is reported as case study by Cafiso et al in press at the next 93rd TRB (January, 2014) [4.4] about the estimation of the safety effects of a new class of barriers meeting the EU Standard, using data of a motorway in Italy. That study was integrated with more data and the calibration of a Crash Modification Function. To do this a cluster analysis was performed on the Curvature Change Rate. The Crash Modification Function give good results for the ran-off-road crashes while for the total crashes the R^2showed poor results.

References

[4.1]. Highway Safety Manual (HSM). First Edition. FHWA American Association of State Highway and Transportation Officials. U.S.A. 2010.

[4.2]. Bonneson. J.A.. Pratt. M.P.. “A Procedure for Developing Accident Modification Factors from Cross Section Data”. Transportation Research Board. National Research Council. Washington. D.C.. 2008.

[4.3]. Frank. G.. Persaud. B.. Lyon. L.. A Guide to Developing Quality Crash Modification Factors. U.S. Department of Transportation and Federal Highway Administration (FHWA). Report No. FHWA-SA-10-032.

[4.4]. Cafiso. S.. D'Agostino. C.. Persaud. B.. Investigating the influence on safety of retrofitting Italian motorways with barriers meeting a new EU standard. In press in Transportation Research Board 93rd Annual Meeting Proceedings. Washington D.C.. 2014.

[4.5]. Hauer. E.. Observational Before-After Studies in Road Safety. Pergamon Press. Elsevier Science Ltd.. Oxford. England. 1997.

[4.6]. Pendleton. O.. "Evaluation of Accident Analysis Methodology." Report FHWA-RD-96-039. Federal Highway Administration. Washington. D.C.. 1996.

[4.7]. Council. F. M.. Persaud. B.. Eccles. K.. Lyon. C.. and Griffith. M. S. (2005). Safety Evaluation of Red-Light Cameras. Battelle Memorial Institute Federal Highway Administration.

[4.8]. Harwood. D. W.. Bauer. K. M.. Potts. I. B.. Torbic. D. J.. Richard. K. R.. Kohlman Rabbani. E. R.. Hauer. E.. and Elefteriadou. L. (2002). "Safety Effectiveness of Intersection Left- and Right-Turn Lanes." FHWA-RD-02-089 Federal Highway Administration. McLean. Va.

[4.9]. Persaud. B.N.. R. Retting. P. Garder. and D. Lord. "Observational Before-After Study of the Safety Effect of U.S. Roundabout Conversions Using the Empirical Bayes Method". Presented at the 80th Annual Meeting of the Transportation Research Board. Washington. D.C.. 2001.

[4.10]. Persaud. B. N.. Retting. R. A.. and Lyon. C. (2003). Crash Reduction Following Installation of Centerline Rumble Strips on Rural Two-Lane Roads. Insurance Institute for Highway Safety.

[4.11]. Persaud. B.. Council. F. M.. Lyon. C.. Eccles. K.. and Griffith. M. (2005). Multijurisdictional Safety Evaluation of Red Light Cameras. Transportation Research Board.

[4.12]. Persaud. B.. Lyon. C.. Empirical Bayes before–after safety studies: Lessons learned from two decades of experience and future directions. Accident Analysis and Prevention 39 (2007) 546–555.

[4.13]. Persaud. B.. Lana B.. Lyon. C.. Bhimb R.. Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. Accident Analysis and Prevention 42 (2010) 38–43.

[4.14]. FHWA American Association of State Highway and Transportation Officials. Safety Analyst: Software Tools for Safety Management of Specific Highway Sites. U.S.A. 2010.

[4.15]. Miaou. S.P. Measuring the Goodness-of-Fit of Accident Prediction Models. FHWA-RD-96-040. Federal Highway Administration. Washington. D.C.. 1996.

[4.16]. NCHRP Report 350. Recommended Procedures for the Safety Performance Evaluation of Highway Features. National Cooperative Highway Research Program - Transportation Research Board National Research Council. U.S.A. 1993.

[4.17]. Bonneson. J.. Lord. D.. Zimmerman. K.. Fitzpatrick. K.. Pratt M.. Development of tools for evaluating the safety implications of highway design decisions. Report FHWA/TX-07/0-4703-4. Austin. Texas. 2006.

[4.18]. Fitzpatrick. k.. Lord. D.. Park. B. J.. Accident Modification Factors for Medians Freeways and Multilane Rural Highways in Texas. Transportation Research Record: Journal of the Transportation Research Board. No. 2083. Transportation Research Board of the National Academies. Washington. D.C.. 2008. pp. 62–71.

[4.19]. Elvik. R.. Hoye. A.. Vaa. T.. Sorensen. M.. Handbook of road safety measure (Second Edition). Emeral Group Bingley. UK. 2009.

[4.20]. Scully. J.. Newstead. S.. Corben. B.. Candappa. N.. 2006. Evaluation of the effectiveness of the $240M statewide blackspot program-accident blackspot component. 2006. Monash University Accident Research Centre. Clayton. Victoria. Australia.

[4.21]. Turner. B.. Imberger. K.. Roper. P.. Pyta. V.. McLean. J.. Road Safety Engineering Risk Assessment Part 6: Crash Reduction Factors. Publication Austroads technical report AP-T151/10. 2010.

[4.22]. Zegeer. C.V.. Hummer. J.. Herf. L.. Reinfurt. D.. Hunter. W.. Safety cost-effectiveness of incremental changes in cross-section design-informational guide. FHWA/RD-87-094.

United States Federal Highway Administration. McLean. VA.1987.

[4.23]. http://www.cmfclearinghouse.org/. Section "Countermeasure: Improve guardrail". Access on July 25. 2013

[4.24]. Gan. A.. Shen. J.. and Rodriguez. A.. Update of Florida Crash Reduction Factors and Countermeasures to Improve the Development of District. Safety Improvement Projects. Florida Department of Transportation. 2005.

[4.25]. Cafiso. S.. D'Agostino. C.. Persaud. B.. Investigating the influence of segmentation in estimating safety performance functions for roadway sections. Transportation Research Board Annual Meeting Proceedings. Washington D.C.. 2013.

[4.26]. I dati sugli incidenti stradali rilevati nel 2002. 2003. 2004. 2005. 2006 2007. 2008 e 2009. Istituto Nazionale di Statistica. Milan. Italy. 2010.

[4.27]. Frank. G.. Persaud. B.. Lyon. L.. A Guide to Developing Quality Crash Modification Factors. U.S. Department of Transportation and Federal Highway Administration (FHWA). Report No. FHWA-SA-10-032.

[4.28]. Lord. D. and Persaud. B.. Accident Prediction Models With and Without Trend. Transportation Research Record: Journal of the Transportation Research Board. No. 1717. Transportation Research Board of the National Academies. Washington. D.C.. 2007. pp. 40–48.

[4.29]. Cafiso. S.. D'Agostino. C.. Safety Performance Function for Motorways using Generalized Estimation Equations. Procedia - Social and Behavioral Sciences Volume 53. 3 October 2012. Pages 901–910.

[4.30]. SAS Help and Documentation. Software of SAS 9.2. Copyright (c) 2002-2008 by SAS Institute Inc.. Cary. NC. USA.

[4.31]. Higle. J. L.. J. M. Witkowski. Bayesian Identification of Hazardous Locations. In Transportation Research Record 1185. TRB. National Research Council. Washington. D.C.. 1988. pp. 24–36.

[4.32]. Miaou. S.-P.. Lord. D.. Modeling Traffic Crash–Flow Relationships for Intersections: Dispersion Parameter. Functional Form. and Bayes Versus Empirical Bayes Methods. Transportation Research Record: Journal of the Transportation Research Board. No. 1840. Transportation Research Board of the National Academies. Washington. D.C.. 2003. pp. 31–40.

[4.33]. Heydecker. B. G.. Wu. J.. Identification of Sites for Road Accident Remedial Work by Bayesian Statistical Methods: An Example of Uncertain Inference. Advances in Engineering Software. Vol. 32. 2001. pp. 859–869.

[4.34]. Hauer. E.. Overdispersion in Modelling Accidents on Road Sections and in Empirical Bayes Estimation. Accident Analysis and Prevention. Vol. 33. No. 6. 2001. pp. 799–808.

[4.35]. Cafiso. S.. Di Silvestro. Persaud. B.. Begum. M.A.. Revisiting Variability of Dispersion Parameter of Safety Performance for Two-Lane Rural Roads. Transportation Research Record: Journal of the Transportation Research Board. No. 2148. Transportation Research Board of the National Academies. Washington. D.C.. 2010. pp. 38–46.

[4.36]. Hauer. E.. and J. Bamfo. Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables. Proc.. ICTCT Conference. Lund. Sweden. 1997.

[4.37]. Hauer, E., Bonneson, J. A., Council, F., Srinivasan, R., and Zegeer, C.. Crash Modification Factors, Foundational Issues. Transportation Research Record: Journal of the Transportation Research Board, No. 2279, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 67–74.

[4.38]. Bailey, Ken. 1994. Typologies and Taxonomies-Numerical Taxonomy and Cluster Analysis (sagepub.com).

[4.39]. Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. Journal of Abnormal and Social Psychology, 38, 476-506.

[4.40]. Estivill-Castro, Vladimir (June 2002 2002). "Why so many clustering algorithms — A Position Paper". ACM SIGKDD Explorations Newsletter 4 (1): 65 75. doi:10.1145/568574.568575.

[4.41]. Wikipedia http://en.wikipedia.org/wiki/Cluster_analysis. Access October 2013.

[4.42]. Statgraphics Statistical Analysis and Data Visualization Software. Software and documentation, version 2013.

[4.43]. http://www.cmfclearinghouse.org/. Access on September 25. 2013.

CHAPTER 5

THE TRADITIONAL TECHNIQUES FOR THE BENEFIT-COST ANALYSIS AND THE CROSS SITE VARIANCE OF A CMF

5.1. Introduction

The best methodology of calibration of CMFs is well known to be based on stochastic approach. The problem of regression to the mean and the selection bias can be controlled using a sophisticated probabilistic approach introduced by Hauer [5.1] in 1997 and developed by various author in the last 2 decades [5.2][5.3][5.4][5.5][5.6][5.7][5.8].

The new methodologies developed for the calibration of the Safety Performance Functions are pushing the Authors to find new advanced methodology able to address the problem of time trend [5.9] and spatial correlation of data and to use more complicated model form and different distribution of outcomes [5.10] in the calibration of CMFs.

Despite the efforts on the calibration of CMFs to improve reliability, evaluation of safety benefits of applying a treatment continue to be performed using a deterministic approach. The traditional techniques for the evaluation of the benefits of a treatment don't take into account the statistical distribution of the CMFs and their stochastic nature.

In the first edition of the Highway Safety Manual [5.11] the methodology for the economical evaluation of a treatment was standardized. The diagram is reported in Figure 5.1. The main issue is the quantification of crash reduction, and the application of the traditional methodology for the economical evaluation of the alternatives based on the CMFs.

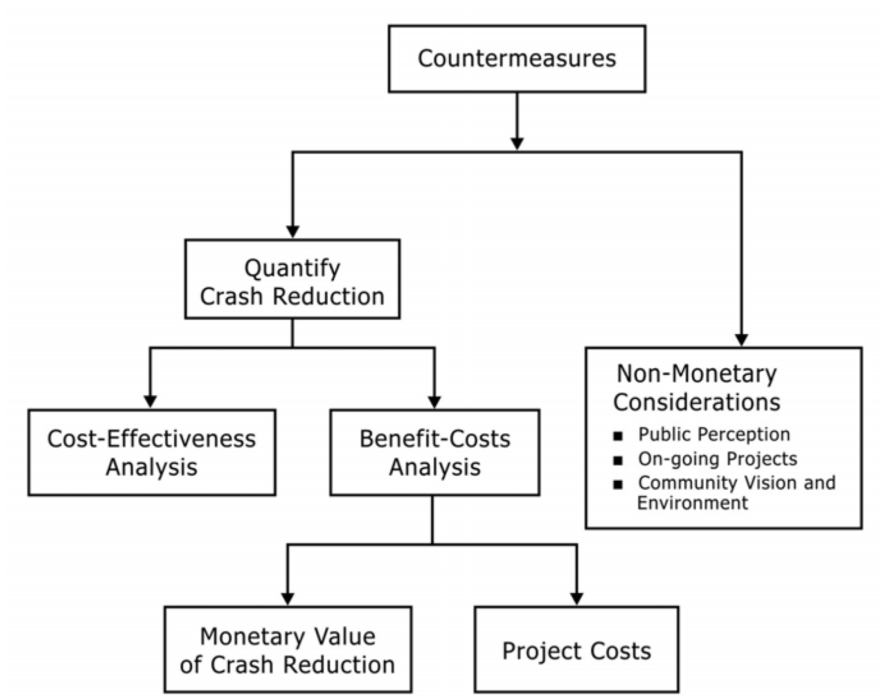


Figure 5.1. Methodology for economical valuation of a countermeasure proposed by HSM.

The new European regulation on Road Safety (96/2008 CE) imposes on each Member State a strict control of the treatment both for new projects and for existing roads, with the aims, and the new concept, to take into account safety in each phase of the design and/or in the construction.

The safety benefit-cost analysis is becoming determinant in the choice of a treatment or of the geometric design of a new road, and it is becoming predominant in the choice of alternatives, or in the evaluation of the final cost of the entire project. In this period of economic trouble the reliability of a countermeasure plays a fundamental role in the decision making process.

In the chapter all the elements that are presented in the different phase of a benefit cost analysis are reported together with the traditional techniques and a new stochastic approach to the problem.

5.2. The cost of crashes

As it is clear from figure 4.1, one of the more important tasks is to quantify the cost of a crash. It depends on the severity of crashes, the geographic location intended as the economic condition of the Country in which the crash took place. In the Highway Safety Manual the reported cost are based on *"Crash Cost Estimates by Maximum Police-Reported Injury Severity within Selected Crash Geometries"* [5.12] in which the cost are evaluated for different crash severity (Table 5.1).

In Italy three level of severity are defined for the crashes cost:

- Fatal;

- Serious injury; and
- Property damage only.

Table 5.1. Crash cost reported by HSM.

Collision type	Comprehensive Crash Costs
Fatality (K)	$4,008,900
Disabling Injury (A)	$216,000
Evident Injury (B)	$79,000
Fatal/Injury (K/A/B)	$158,200
Possible Injury (C)	$44,900
PDO (O)	$7,400

The economical evaluation of those categories is reported on Table 5.2 and the source is the "*Social crash cost estimation (2012)*" from the Ministry of Transportation [5.13].

Table 5.2. Crash cost reported by Italian Ministry of Transportation.

Collision type	Social cost referred to 1 person
Fatal	€ 1,503,990.00
Serious injury	€ 42,219.00
Property damage only	€ 7,686.00

Those cost reported on Table 5.2 are not referred to as crashes but as a single person involved in a crash. However often is not feasible to know the number of persons involved and, in that case the total cost of a crash, have to be used. For this reason the percentage of the number of persons involved in a crash is reported on table 5.3 in comparison with the number of crashes for road categories in Italy.

Table 5.3. Percentage of number dead or injured person for road categories for crashes.

Area/Road	Fatal./(100 Crashes)	Inj./(100 Crashes)	Crash cost (Italy-2002)
Motorway	5.30%	173.81%	€ 132.051.74
Rural	6.97%	159.33%	€ 151.785.30
Urban	1.68%	135.53%	€ 67.708.10

In the "*Social crash cost estimation (2012)*" from the Ministry of Transportation [5.13], is reported a comprehensive cost of a crash related to the severity. In table 5.4 the comprehensive cost for crash severity is reported.

Table 5.4. Comprehensive average cost for crash severity reported by Italian Ministry of Transportation.

Severity	Comprehensive average cost
Fatal	€ 1,642,236.00
Injury	€ 309,863.00

Hauer et al. [5.14] used an average freeway crash cost, including all the categories for nighttime crashes of $20,000.00 in a study extensive reported later in the Chapter.

5.3. Traditional techniques for the benefit-cost analysis

In the first edition of the Highway Safety Manual [5.11] three different methodologies are reported for the benefit-cost analysis:

- Net Present Value (NPV);

- Benefit-Cost Ratio (BCR); and
- Cost-Effectiveness Index.

In the following the three methodologies are described with a detailed explanation of the element or terminology used in the elaboration.

Table 5.4. Data Needs for Calculating Project Benefits reported by HSM.

Activity	Data Need
Calculate Monetary Benefit	
Estimate change in crashes by severity	Crash history by severity
	Current and future Average Annual Daily Traffic (AADT) volumes
	Implementation year for expected countermeasure
	SPF for current and future site conditions (if necessary)
	CMFs for all countermeasures under consideration
Convert change in crash frequency to annual Monetary value of crashes by severity	monetary value Change in crash frequency estimates
	Service life of the countermeasure
Convert annual monetary value to a present value	Discount rate (minimum rate of return)
Calculate Costs	
Calculate construction and other implementation costs	Subject to standards for the jurisdiction
Convert costs to present value	Service life of the countermeasure(s)
	Project phasing schedule

To calculate the benefit of a countermeasure a serious of data need before the conversion in monetary value.

For all the methods described above, some parameters need to be estimated to obtain a monetary value of the benefit. In Table 5.4 a series of activity and data that need to calculate. The information reported in table 5.4 are taken from the HSM.

5.3.1. The net present value (NPV)

The net present value (NPV) method is also referred to as the net present worth (NPW) method. This method is used to express the difference between discounted costs and discounted benefits of an individual improvement project in a single amount. The term "discount" indicates that the monetary costs and benefits are converted to a present value using a discount rate.

The discount rate is an interest rate that is chosen to reflect the time value of money. The discount rate represents the minimum rate of return that would be considered by an agency to provide an attractive investment. Thus, the minimum attractive rate of return is judged in comparison with other opportunities to invest public funds wisely to obtain improvements that benefit the public. Two basic factors to consider when selecting a discount rate:

a) The discount rate corresponds to the treatment of inflation (i.e., real dollars versus nominal dollars) in the analysis being conducted. If benefits and costs are estimated in real (uninflated) dollars, then a real discount rate is used. If benefits and costs are estimated in nominal (inflated) dollars, then a nominal discount rate is used.

b) The discount rate reflects the private cost of capital instead of the public sector borrowing rate. Reflecting the private cost of capital implicitly accounts for the element of risk in the investment. Risk in the investment corresponds to the potential that the benefits and costs associated with the project are not realized within the given service life of the project.

Discount rates are used for the calculation of benefits and costs for all improvement projects. Therefore, it is reasonable that jurisdictions are familiar with the discount rates commonly used and accepted for roadway improvements. Further guidance is found in the American Associate of State Highway and Transportation Officials (AASHTO) publication entitled A Manual of User Benefit Analysis for Highways (also known as the AASHTO Redbook) [5.15].

The NPV method is used for the two basic functions listed below:

- Determine which countermeasure or set of countermeasures provides the most cost-efficient means to reduce crashes. Countermeasure(s) are ordered from the highest to lowest NPV.
- Evaluate if an individual project is economically justified. A project with a NPV greater than zero indicates a project with

benefits that are sufficient enough to justify implementation of the countermeasure.

To apply the methodology of NPV need to estimate the crash reduction due to the countermeasure using the specific CMF and convert the change in estimated average crash frequency to an annual monetary value to representative of the benefits. The benefit calculated in this way are referred to the moment of the implementation of the countermeasure, and they don't take into account the benefit in the years after, in the whole service life of the treatment.

All improvement projects have a service life. In terms of a countermeasure, the service life corresponds to the number of years in which the countermeasure is expected to have a noticeable and quantifiable effect on the crash occurrence at the site. Some countermeasures, such as pavement markings, deteriorate as time passes, and need to be renewed. For other countermeasures, other roadway design modifications and changes in the surrounding land uses that occur as time passes may influence the crash occurrence at the site, reducing the effectiveness of the countermeasure. The service life of a countermeasure reflects a reasonable time period in which roadway characteristics and traffic patterns are expected to remain relatively stable.

When the annual benefits are uniform over the service life of the project Equations 5.1 and 5.2 can be used to calculate present value of project benefits.

$$PV_{benefits} = Total\ Annual\ Monetary\ Benefits\ x\ (P/A, i, y) \quad (5.1)$$

where,

- $PV_{benefits}$ = Present value of the treatment benefits for a specific site, v; and
- (P/A, I, y) = conversion factor for a series of uniform annual amounts to present value

$$\left(P/A,i,y\right)=\frac{\left(1+i\right)^{y}-1}{i\cdot\left(1+i\right)^{y}} \tag{5.2}$$

i= Minimum attractive rate of return or discount rate; and

y= Year in the service life of the countermeasure(s)

Some countermeasures yield larger changes in expected average crash frequency in the first years after implementation than in subsequent years. In order to account for this occurrence over the service life of the countermeasure, non-uniform annual monetary values can be calculated considering separately each benefit and summing overall in the year of service life.

Than is possible to calculate the present value of the benefit at the generic year y as follow:

$$Benefit=\sum_{i=0}^{N}\frac{Benefit_{i}}{(1+i)^{y}} \tag{5.3}$$

In the same way the cost of the treatment has to be evaluated considering the present value.

To calculate the Net Present Value the following equation has to be applied:

$$NPV = PV_{Benefits} - PV_{Costs} \tag{5.4}$$

where,

$PV_{Benefits}$= Present value of the countermeasure benefits; and

PV_{Costs}= Present value of the countermeasure costs.

If the NPV > 0 then the individual treatment is economically justified.

5.3.2. Benefit-Cost Ratio (BCR)

A benefit-cost ratio is the ratio of the present-value benefits of a project to the implementation costs of the project (BCR = Benefits/Costs). If the ratio is greater than 1.0, then the project is considered economically justified. Countermeasures are ranked from highest to lowest BCR. An incremental benefit-cost analysis is needed to use the BCR as a tool for comparing project alternatives.

This method is used to determine the most valuable countermeasure(s) for a specific site and is used to evaluate economic justification of individual projects. The benefit-cost ratio method is not valid for prioritizing multiple projects or multiple alternatives for a single project; the methods discussed in Chapter 8 are valid processes to prioritize multiple projects or multiple alternatives.

To apply the method of the benefit cost ratio need to estimate the average crash frequency of the sites, and the present value of the costs associated with the safety improvement treatment and calculate the ratio:

$$BCR = \frac{PV_{Benefits}}{PV_{Costs}} \tag{5.5}$$

where,

BCR = Benefit cost ratio;

$PV_{Benefits}$ = Present value of project benefits; and

PV_{Cost} = Present value of project costs.

If the BCR is greater than 1.0, then the treatment is economically justified.

5.3.3. Cost-Effectiveness Index

In cost-effectiveness analysis the predicted change in average crash frequency are not quantified as monetary values, but are compared directly to project costs. The cost-effectiveness of a countermeasure implementation project is expressed as the annual cost per crash reduced. Both the project cost and the estimated average crash frequency reduced must apply to the same time period, either on an annual basis or over the entire life of the project. This method requires an estimate of the change in crashes and cost estimate associated with implementing the countermeasure. However, the change in estimated crash frequency is not converted to a monetary value.

This method is used to gain a quantifiable understanding of the value of implementing an individual countermeasure or multiple countermeasures at an individual site when an agency does not support the monetary crash cost values used to convert a project's change in estimated average crash frequency reduction to a monetary value.

To apply the method of the benefit cost ratio need to estimate the change in expected average crash frequency due to the safety improvement treatment and calculate the costs associated with implementing the treatment and Calculate the cost-effectiveness of the safety improvement project at the site by dividing the present value of the costs by the estimated change in average crash frequency over the life of the countermeasure with the following equation:

$$Cost - Effectiveness\ Index = \frac{PV_{Costs}}{N_{Predicted} - N_{Observed}} \quad (5.6)$$

where,

PV_{Costs} = Present value of the treatment cost;

$N_{Predicted}$ = Predicted crash frequency for year y; and

$N_{Observed}$ = Observed crash frequency for year y.

5.4. The cross site variance of the CMF and CMFunction

Hauer argued logically that two different CMFs calibrated for the same treatment, used an appropriate methodology, but in different sites or region, have to be different for a reason that can be investigated with more research on the topic [5.16]. There is no reason to think that one of the CMFs is affected by some bias but at the same time the use of the CMF is not univocal, because the different values assumed in the different estimations. The same happens when a cross site distribution is considered, with the difference that the variance of the CMFs calibrated on a single site is larger in comparison with the variance calibrated on the whole sample.

The variance of the CMF estimated using an empirical Bayes before after analysis, through the Equation 4.14, is strongly influenced by the sample size. It is an indicator of the reliability of CMF calibration and of the sample size used for the elaboration.

The combination of more CMFs related to the same treatment proposed by Hauer et al. [5.16], is based on a consideration that a CMF calibrated on a sample is an estimate of the long term value of the reduction of crashes due to the treatment investigated and that this estimation is influenced by a factor logically derived by the location in which the CMF is estimated or to a common geometric feature present in the sites used for the calibration. An extensive discussion about Hauer et al [5.14][5.16] researches on the topic is reported below and in the next Paragraph, using also some verbatim expression of the Authors that can help in the explanation of the methodology.

The Agencies have to make the decision to implement a treatment based on the crash reduction of the treatment itself.

The comparison is always between the expected target crash frequency of the action implemented, denoted by μ_a, and the expected target crash frequency prevailing under identical conditions but without the action having been implemented, denoted by μ_b. Research results usually report estimates of the ratio μ_a divided by μ_b. This ratio is the CMF of implementing a instead of b, to be denoted by θ(a;b), or, by θ. Thus

$$CMF = \theta(a;b) = \frac{expected\ crashes\ with\ a}{expected\ crashes\ with\ b} = \frac{\mu_a}{\mu_b} \tag{5.7}$$

When the implementation of a instead of b reduces the expected target crash frequency, then θ(a;b) < 1. The main use of the estimate $\hat{\theta}$(a;b) of θ(a;b) is to predict what is expected to be the safety effect of doing *a* instead of *b* in some specific circumstance. Transposing the terms in Equation 5.7 and adding the caret to signify “estimate” gives:

$$\hat{\mu}_a = \hat{\mu}_b \cdot \hat{\theta}(a;b) \tag{5.8}$$

The safety effect of implementing *a* instead of *b* is usually measured by the expected change in the number of target crashes (by severity). The estimate of this expected change is:

$$\hat{\mu}_b - \hat{\mu}_a = \hat{\mu}_b \cdot \left[1 - \hat{\theta}(a;b)\right] \tag{5.9}$$

Clearly, $\hat{\theta}$ (a, b) is needed to predict the safety effect.

There is no reason to believe that a CMF has the same effect on safety everywhere and at all times. The effect may depend on the specifics of the treatment, and the feature of the roads, of the road users, the traffic volume and so forth. That is, it should not be assumed that θ(a;b) is a universal constant that has the same value always and everywhere. Rather, θ(a;b) should be viewed as a random variable, the value of which depends on a host of factors. These factors, taken together, are referred by Hauer at al. [5.16] as the circumstances of implementation.

Since θ(a;b) is a random variable, it has a probability distribution with a mean and a variance. For some actions, the θ(a;b) may vary little from one implementation to another, and therefore the variance will be small; for other actions the variance may be large.

Hauer et al. [5.16] [5.17] logically considered the distribution of θ(a,b) as a factor influencing the transferability of the CMF.

Thinking of θ as a random variable allows the question of transferability to be correctly framed. The issue is: in a cost effectiveness or cost–benefit framework, decisions are based on expected consequences. This is why, to predict the future safety effect with Equation 5.8, $\bar{\theta}$ is used, in other words, the current estimate of the expected value of θ based on past research.

The difference between θ and $\bar{\theta}$ determines whether the decision about implementing a treatment is right or wrong.

Thus, concern about transferability amounts to concern about how well the $\bar{\theta}$ based on past implementations predicts the θ of a future implementation. When past research indicates that whenever *a* was implemented instead of *b* approximately the same θ was found, the issue of transferability should not arise. Transferability concerns are real when the difference between $\bar{\theta}$ and θ is frequently large.

Thus, concern about transferability arises whenever the variance of θ is large or when $\bar{\theta}$ is not a good estimate of the mean θ; it arises irrespective of whether the future application is in a different country, city, project, or time period.

The setting is that of making a decision about a future action that has safety consequences. For that purpose, the θ of that future action will need to be predicted. The assumption is that the future will be similar to the past. If so, the θ for the future action will be one of the values from the probability distribution of past θs, the standard deviation of which is σ{θ}. For decision making, it is best to assume that the θ of the future action is the current estimate $\bar{\theta}$ of E{θ}, the mean of past experiences. The $\bar{\theta}$ has a standard error to be denoted by s{$\bar{\theta}$}. The σ{θ} and s{$\bar{\theta}$} are two different constructs.

While σ{θ} is an aspect of reality—namely, how variable the CMFs are from one circumstance to another—the s{$\bar{\theta}$} measures the uncertainty of an estimate and thus, indirectly, the quality of data. If s{$\bar{\theta}$} and σ{θ} are small, then Equation 5.8 can be confidently used to predict the safety effect of implementing *a* instead of *b*. If s{$\bar{\theta}$} or σ{θ} are large, then predictions made with Equation 5.8 can be insufficiently accurate. A prediction of θ is insufficiently accurate when some likely-to-occur values of θ lead to the decision to implement and other likely-to-occur values lead to the opposite decision. Decisions based on insufficiently accurate predictions are in danger of being wrong. It follows that rational decision making about actions that have safety consequences requires three estimates: the current estimate $\bar{\theta}$ of E{θ}, its standard error s{$\bar{\theta}$}, and an estimate of σ{θ}. Whereas the decision to implement or not implement is based on $\bar{\theta}$, both s{$\bar{\theta}$} and σ{θ} are needed to know whether the decision can be made with confidence.

The usual sources of CMFs are the Highway Safety Manual [5.11], the Handbook of Road Safety Measures [5.18], and FHWA's

Crash Modification Factor Clearinghouse website [5.19]. These publications list $\bar{\theta}$s and sometimes their s{$\bar{\theta}$}. None give estimates of σ{θ}.

To estimate $\bar{\theta}$ and Its s{$\bar{\theta}$} two methods are preferred, both described by Hauer [5.1]. One is a simple estimate of $\bar{\theta}$ and s{$\bar{\theta}$} with a weighted average of the $\hat{\theta}$s using as weight their own standard error. Thus

$$\overline{\theta} = \sum_{i=1}^{n} \frac{\frac{1}{s_i^2}}{\sum_{i=1}^{n} \frac{1}{s_i^2}} \cdot \hat{\theta}_i \tag{5.10}$$

$$Var\left\{\overline{\theta}\right\} = s^2\left\{\overline{\theta}\right\} = \frac{1}{\sum_{i=1}^{n} \frac{1}{s_i^2}} \tag{5.11}$$

A second methodology, if more information is available on CMFs is to estimate $\bar{\theta}$, and Its s{$\bar{\theta}$} using the Equation 5.13 and the 5.14 respectively.

In Chapter 4, the $\bar{\theta}$s and their s{$\bar{\theta}$} were estimated using the cluster analysis for different sites and for different value of CCR using equation 4.13 and 4.14. Table 4.15 reported the results. The average weighted value do not need to be estimated because the CMF is calibrated on the same data on the same sites. In other terms the best estimation of $\bar{\theta}$ in the case of the calibration of a CMF, when no other CMFs for the same treatment are available in literature, is θ. The results are shown in the Table 5.11 and are related to the CMF (θ) calibrated on the whole sample.

If it could be assumed that the considered CMF is the same in each site, then the variance calculated in that way would describe the uncertainty of the CMF. However, such an assumption would be unreasonable. The uncertainty about the application of that treatment is not only due to $\bar{\theta}$; it is also caused by the question of how variable θ is from one site to another. This variability is measured by σ{θ}. Thus, the next task is to use the available data to estimate σ{θ}. The advantage to considering a cross site distribution of the CMF is the possibility to estimate the cross site variance that should be used in the benefit/cost analysis.

To estimate the cross site variance Hauer et al. [5.16] consider the law of total variance a law that follows by logic from the axioms of probability. With this law, it can be shown that:

$$Var\{\theta\} = Var\{\hat{\theta}\} - E\left\{Var\{\hat{\theta}|\theta\}\right\} \quad (5.12)$$

So considering the cross site variance as the difference of the variance of the $\hat{\theta}$ calculated in the Chapter 4 and reported in the Table 4.13 minus the covariance of the $\hat{\theta}s$ calculated in different sites.

The total variance of the CMF, $Var^{*}\{\theta\}$, is defined as the sum of the $\hat{\theta}s$ and a factor that describes the variance of the cross site CMF.

Thus:

$$Var^{*}\{\theta\} = \hat{V} + Var\{\bar{\theta}\} \quad (5.13)$$

Where:

$$\hat{V} = \begin{cases} \frac{\sum_{i=1}^{n}(\hat{\theta}_i - \bar{\theta})^2}{n} - \frac{\sum_{i=1}^{n} s_i^2}{n} & if\ positive \\ 0 & otherwise \end{cases} \quad (5.14)$$

In the first edition of the Highway Safety Manual, $\hat{V}$ is not used to describe the uncertainty about θ . In this illustration, the Highway Safety Manual would list a variance that is about one third of the correct value [5.16].

Thus two main issue arise from the Hauer et al. study [5.16], one is the variance of the CMF plays a fundamental role in the decision making process when a cross sites distribution of the CMF (or a cross CMFs distribution) is considered. The second is that the simple variance of the CMFs alone is not able to assess the reliability or the transferability of the CMFs themselves but a cross site variance (or a cross CMFs variance) need to be estimated to implement a reliable decision making process.

If the procedure proposed by Hauer is applied on the results of the cluster analysis reported on Chapter 4, it is possible to compute the variance for each cluster and the total cross site variance of the CMF. In table 5.5 the variance is reported for the total and the ran off road crashes for each cluster.

Thus through the Equation 5.13 is possible to combine the variance of the CMF of each cluster to obtain a single value of the cross site variance of the CMF. Using the unbiased mean to compute the value of $\hat{V}$, as suggested by Hauer [5.16] with the

Equation 5.14 and considering the Equation 5.13 the results are shown below.

Table 5.5. Cross site variance of the CMF calculated for the cluster analysis for the total crashes and the ran-off-road crashes.

Total					
Cluster (CCR)	**Length (m)**	**Observed**	**CMF**	**Variance**	**Cluster**
1.1082	2823.495	10	0.65	0.404	1
27.3214	7798.79	23	0.49	0.108	2
58.7093	2991.85	28	1.28	0.210	3
103.038	3079.73	8	0.56	0.277	4 - 5
Ran-off-Road					
Cluster (CCR)	**Length (m)**	**Observed**	**CMF**	**Variance**	**Cluster**
1.1082	2823.495	6	0.33	0.350	1
27.3214	7798.79	10	0.32	0.097	2
58.7093	2991.85	4	0.26	0.100	3
103.038	3079.73	2	0.14	0.132	4 - 5

$$\hat{V}_{Total} = 0.018 \quad (5.15)$$

$$\hat{V}_{Ran-off-Road} = 0.011 \quad (5.16)$$

The cross site variance for the considering CMFunctions are computed with Equation 4.13 for the total crashes:

$$Var^{*}\{\theta\} = \hat{V} + Var\{\bar{\theta}\} = 0.018 + 0.09 = 0.027 \quad (5.17)$$

And for the ran-off-road-crashes

$$Var^{*}\{\theta\} = \hat{V} + Var\{\bar{\theta}\} = 0.011 + 0.04 = 0.016 \quad (5.18)$$

As it is clear from the Equations 5.17 and 5.18 the cross site variance of the CMF give a contribution to the overall variance that is double respect to the simple CMF variance.

The procedure presented above to evaluate the cross site variance has a limitation in this application. The cross site variance has to be able to describe how the CMF varying from a site to another. In other terms it should be able to catch the different conditions (related e.g. with road feature, traffic composition, geometry, etc.) present in different segments. When it is applied on the cluster analysis the value of the cross site variance is underestimated or overestimated, because the variability related to one of the more important variables, which influence the target crashes, the curvature in that case, is controlled by the cluster analysis.

Furthermore a great variability form a site to another in this case is given by the different lengths of segments. To eliminate the influence of the length of segments a fixed length segmentation approach is suggested.

5.5. Chapter summary

In the Chapter 5 the traditional techniques for the benefit-cost analysis are reported. The great part of the reference of that part is taken by Highway Safety Manual [5.11]. At the end of the Chapter a methodology developed by Hauer et al. [5.16] about the CMF transferability and the evaluation of the cross site distribution of CMF is reported and applied on the results of the cluster analysis from the Chapter 4. The methodology developed by Hauer et al. is referred to different CMF related to the same treatment, but developed in different location. The applied methodology is referred to the same treatment as well but on the dataset used to the estimation of the Crash Modification Function. Indeed the treatment has different effects in each site, and this variability could be used to compute a “cross site variance” of the CMF. In other terms, the treatment could be different effect in different site for reason that cannot be investigated with the data available. This variability could be used to introduce a new approach to the benefit cost analysis that will be described later in the Chapter 6. Although the methodology applied on the Cluster analysis compute on Chapter 4 in general overestimate or underestimate the cross site variance, because the CCR variability is mostly controlled by the cluster analysis, that variability could be still compute using the methodology described in the Chapter.

References

[5.1]. Hauer. E.. Observational Before-After Studies in Road Safety. Pergamon Press. Elsevier Science Ltd.. Oxford. England. 1997.

[5.2]. Frank. G.. Persaud. B.. Lyon. L.. A Guide to Developing Quality Crash Modification Factors. U.S. Department of Transportation and Federal Highway Administration (FHWA). Report No. FHWA-SA-10-032.

[5.3]. Council. F. M.. Persaud. B.. Eccles. K.. Lyon. C.. and Griffith. M. S. (2005). Safety Evaluation of Red-Light Cameras. Battelle Memorial Institute Federal Highway Administration.

[5.4]. Bonneson. J.A.. Pratt. M.P.. “A Procedure for Developing Accident Modification Factors from Cross Section Data”.

Transportation Research Board. National Research Council. Washington. D.C.. 2008.

[5.5]. Harwood. D. W.. Bauer. K. M.. Potts. I. B.. Torbic. D. J.. Richard. K. R.. Kohlman Rabbani. E. R.. Hauer. E.. and Elefteriadou. L. (2002). "Safety Effectiveness of Intersection Left- and Right-Turn Lanes." FHWA-RD-02-089 Federal Highway Administration. McLean. Va.

[5.6]. Persaud. B.N.. R. Retting. P. Garder. and D. Lord. “Observational Before-After Study of the Safety Effect of U.S. Roundabout Conversions Using the Empirical Bayes Method”. Presented at the 80th Annual Meeting of the Transportation Research Board. Washington. D.C.. 2001.

[5.7]. Persaud. B. N.. Retting. R. A.. and Lyon. C. (2003). Crash Reduction Following Installation of Centerline Rumble Strips on Rural Two-Lane Roads. Insurance Institute for Highway Safety.

[5.8]. Persaud. B.. Lyon. C.. Empirical Bayes before–after safety studies: Lessons learned from two decades of experience and future directions. Accident Analysis and Prevention 39 (2007) 546–555.

[5.9]. Cafiso, S., D’Agostino, C.. Safety Performance Function for Motorways using Generalized Estimation Equations Original Research Article Procedia - Social and Behavioral Sciences, Volume 53, 3 October 2012, Pages 900-909.

[5.10]. Persaud. B.. Lana B.. Lyon. C.. Bhimb R.. Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. Accident Analysis and Prevention 42 (2010) 38–43.

[5.11]. Highway Safety Manual (HSM). First Edition. FHWA American Association of State Highway and Transportation Officials. U.S.A. 2010.

[5.12]. FHWA. Crash Cost Estimates by Maximum Police-Reported Injury Severity within Selected Crash Geometries" - HRT - 05-051, October 2005.

[5.13]. Ministero delle Infrastrutture e dei Trasporti. Direzione Generale per la Sicurezza Stradale. Studio di valutazione dei Costi Sociali dell'incidentalità stradale. 2012.

[5.14]. Hauer, E., Bonneson, J. A., Council, F., Srinivasan, R., and Bahar, G.. Value of Research on Safety Effects of Actions. Transportation Research Record: Journal of the Transportation Research Board, No. 2280, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 68–74.

[5.15]. American Association of State Highway and Transportation Officials (AASHTO). A Manual of User and Non-User Benefit Analysis for Highways. 2010

[5.16]. Hauer, E., Bonneson, J. A., Council, F., Srinivasan, R., and Zegeer, C.. Crash Modification Factors, Foundational Issues. Transportation Research Record: Journal of the Transportation Research Board, No. 2279, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 67–74.

[5.17]. Hauer, E. Accident Modification Functions in Road Safety. Proc., 28th Annual Conference of the Canadian Society for Civil Engineering, London, Ontario, Canada, 2000.

[5.18]. Elvik. R.. Hoye. A.. Vaa. T.. Sorensen. M.. Handbook of road safety measure (Second Edition). Emeral Group Bingley. UK. 2009.

[5.19]. Crash Modification Factor Clearinghouse. FHWA, U.S. Department of Transportation. http://www.cmfclearinghouse.org/. Accessed, September 25th, 2013.

CHAPTER 6

DEVELOPMENT OF A NEW STOCHASTIC APPROACH TO THE BENEFIT-COST ANALYSIS

6.1. Introduction

Considering the cross site variance introduced in the previous Chapter 5 it is possible to think to the CMF as a random variable, with mean the expected value of the CMF and variance the cross site variance introduced above. In this way the benefit cost analysis is strongly influenced by the variance and the results can vary considerably from the deterministic approach. As larger is the variance of the CMF the larger is that difference.

A different methodology is introduced for both existing road and new infrastructure. The difference is related to the variables considered in the calculation of the Benefit-Cost function. Indeed for existing road the EB correction is applied to calculate the expected number of crashes before the treatment is applied, in the case of new infrastructure that value could be considered as a random variable as well with a Gamma distribution. At the end of the Chapter a Montecarlo Simulation is performed to consider a combination of two CMFs together with a frequency analysis to compute the final distribution.

6.2. A stochastic approach to the Benefit-Cost analysis

The basic idea is to consider the variance of the CMF in a benefit cost analysis, or better as the variance could influence the decision making process regarding the chance to implement a treatment or a different combination of them. In order to accomplish this, the methodology introduced by Hauer et al. [6.1][6.2][6.3] was used. Hauer argued logically that not only the variance of the CMF is important but the cross site distribution of the CMF is important for transferability. The greater the cross site variance of a CMF, the higher the probability that the decision to implement or to not implement a treatment is wrong. He introduced a methodology to combine different CMFs related to the same treatment weighing the contribution to the mean to get a unique value obtained using their own variance. The results led to the conclusion that more research on the CMF field can reduce the uncertainty (cfr. Paragraph 5.4).

The second step of his research took into account the quantification of the amount of money that the uncertainty due to a large value of the CMF variance can produce, indicating that as an expected loss. A possible extension of that research could be to consider the expected loss in the decision making process introducing the expected benefits together with the expected loss.

To do that a more detailed explanation of the methodology needs to be done.

The first step is to be clear about what decisions might be improved by taking into account the cross site variance of a CMF. In particular the chance to implement a treatment, or to not implement it, is investigating in both cases using the “expected” benefits and loss that became the targets of the proposed methodology. Decisions of this kind should depend on the balance between the expected benefit and cost. The decision is implemented when:

$$\frac{expected\ benefit}{expected\ cost} > r \quad (6.1)$$

where r is the smallest acceptable benefit–cost ratio and reflects the opportunity cost of the capital available for safety improvements.

It is worth to explain the meaning of r and its implication for the following elaboration. If a deterministic approach is used in the Benefit Cost analysis, the value of r for each treatment and for each segment is a priori determined. In this way, the cost of the treatment in a unit and the benefit are related to the treatment.

For road safety actions, the benefit is the annual value of accident reduction:

$$CMF = \mu(1-\theta)a \quad (6.2)$$

where:

- μ = expected annual number of target crashes on the unit,
- θ = crash modification factor (or function), and

- a = euro value of an average target crash.

There are no reasons to think that the CMF related to the treatment to be applied is not the real value that will occur after the implementation if its variance is not taken into account. However it is prudent to think of θ as a random variable with probability density function (PDF) f(θ), mean E{θ}, and standard deviation σ{θ} [6.4] [6.5]. The PDF f(θ) is assumed to be approximated by the Gamma PDF:

$$f(\theta) = \frac{\theta^{\alpha-1} e^{-\vartheta/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \qquad (6.3)$$

where

$$\alpha = \left(\frac{E\{\theta\}}{\sigma\{\theta\}}\right)^2 \qquad (6.4)$$

$$\beta = \frac{(\sigma\{\theta\})^2}{E\{\theta\}} \qquad (6.5)$$

Of special interest is the θ dividing the implement and the do-not-implement decisions, the break-even θ, denoted θ_{BE}. When $\theta = \theta_{BE}$, the ratio of benefits and costs equals r. Let c denote the annual cost of implementing some action on a unit. From $\mu(1 - \theta_{BE})$ a/c = r, it follows that

$$\theta_{BE} = 1 - \frac{rc}{\mu a} = 1 - \frac{annual\ opportunity\ cost\ of\ implementing\ the\ action}{annual\ cost\ of\ target\ crashes} \qquad (6.6)$$

When $E\{\theta\} < \theta_{BE}$, the decision is implement; it is the wrong decision if $\theta > \theta_{BE}$. When $E\{\theta\} > \theta_{BE}$, the decision is do not implement; it is the wrong decision if $\theta < \theta_{BE}$.

The Hauer et al. research [6.6] was finalized to research the effect of the treatment and the probability that the treatment could not have the expected reduction in the number of target crashes and/or in the severity of target crashes in the optic of transferability.

The following is an important index introduced by Hauer [6.1][6.2]:

$$D = \frac{\left| E\{\theta\} - \theta_{BE} \right|}{\sigma\{\theta\}} \quad (6.7)$$

The larger that D is, the smaller the probability is that decisions will be incorrect. When D is greater than about 2 or 3, this probability is quite small.

That is why D is useful for the identification of research that is of little value, irrespective of any other consideration. To compute the value of D for some action and unit, one needs to have estimates of $E\{\theta\}$, $\sigma\{\theta\}$, and θ_{BE}. For D to be large the denominator of the Equation 6.7 must be small and the numerator must be large. The denominator is small when the safety effect can be accurately predicted, and more research about such actions serves little purpose. The numerator is large when θ_{BE} is either far to the left of $E\{\theta\}$, or far to the right of $E\{\theta\}$.

As can be gleaned from the mathematical expression of θ_{BE}, the larger the cost of implementation of the action is and the smaller the cost of the target crashes on the unit ($\mu \times a$) is, the farther to the left θ_{BE} will be.

Hauer [6.1][6.2] introduced the index D to consider the value of the research, or better how much the new research on the topic can reduce the uncertainty in the application of CMFs considering only the possible loss in the decision of implementing a treatment and the possibility of not implementing a treatment as well.

The proposed methodology takes into account the benefit in both cases. The benefit and the loss in the two possible conditions have to be seen as the Agency perspective. In the case of not implementing the treatment the expected loss are considered by Hauer as the possible loss of money (in a benefit cost analysis) that an Agency could have if the real value of the CMF is smaller than the Break-even point. From an analytical point of view the following expression represents the expected loss:

$$\text{expected loss} = \int_0^{\vartheta_{BE}} L(\theta) f(\theta) d\theta \qquad (6.8)$$

where $L(\theta) = \mu a(\theta - \theta_{BE})$ (6.9)

From this point of view the expected benefit, in the do-not-implement case has to be seen as the right part of the curves (from the break-even point to + ∞). The ratio between the two part of the curves (left and right of the breakeven point respectively is smaller than r indeed. When the θ is smaller than the break-even point the decision should be to implement the treatment. In this case the expected loss introduced by Hauer considered the probability that the CMF used could be bigger than the expected value.

The expression used is the following:

$$\exp ected \quad loss = \int_{\vartheta}^{+\infty} L(\theta) f(\theta) d\theta \tag{6.10}$$

where L(θ) $= \mu a(\theta_{BE} - \vartheta)$ (6.11)

The extension to the methodology proposed by Hauer et al. is to consider the whole PDF function of the CMF, accounting together with the expected loss the expected benefit due to the countermeasure.

In the following figure (Figure 6.1) an example for the first condition ($E\{\theta\} < \theta_{BE}$) when the variance of E{θ} is considered as the variance of the sample (in the example θ =0.33 and σ(θ) = 0.13).

The benefits part (B) of the curves in Figure 6.2 is related to all the values of θ that are smaller than the E{θ}. The neutral part is related to an "acceptable" part. If θ assume a value between E{θ} and θ_{BE}, it is still a benefit from the θ_{BE} point of view, because the Agency accepts each value of θ less than θ_{BE}, but it is a loss at the same time, because the greater value than the expected CMF, for this reason it cannot be considered at the same time both as a benefit, and as a loss.

The real loss is evident when the value of θ is bigger than θ_{BE}. In this cases the decision to implement the treatment is wrong, and the loss could be accounted by the area under the curves that give the probability that θ is bigger than θ_{BE}. This part of the curves can be divided into two different parts.

The first between θ_{BE} and 1 still gives some benefits because the CMF reduces the number of the expected crashes, also if they are not what is expected from an economical point of view. The second part of the curves, called C_2 gives a real cost, both in economic analysis and in terms of crash reduction, because if θ assume a value greater than 1 an increasing number of crashes will be expected.

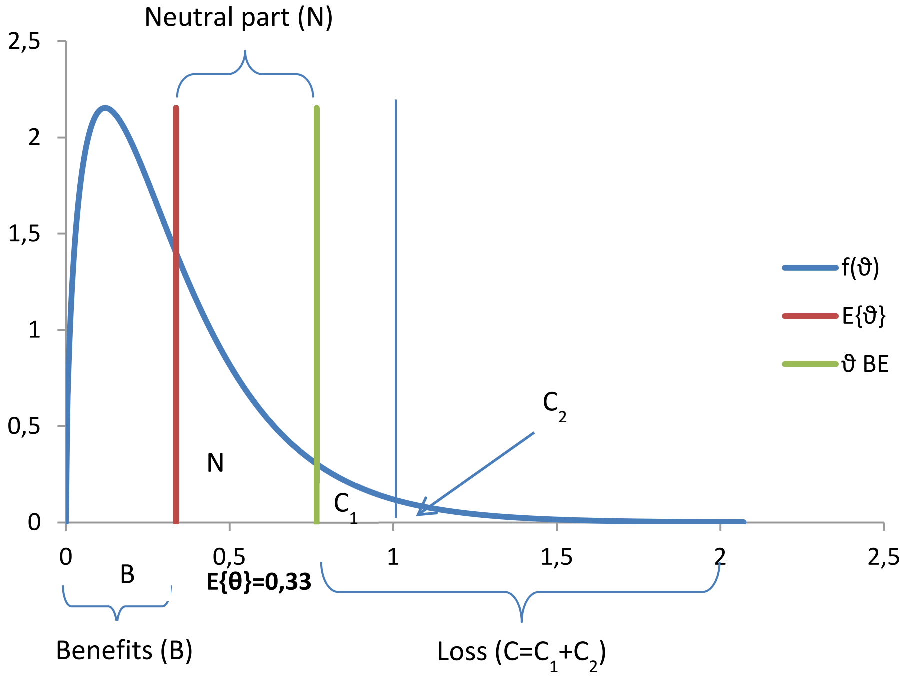


Figure 6.1. The PDF of the f(θ) with the θ_{BE} and the E{θ} (the E{θ} is the red line)

To calculate the benefit and the loss respectively, considering the real distribution of the CMF the following integrals have to be calculate:

$$B = \int_0^{E\{\vartheta\}} L(\theta) f(\theta) d\theta \tag{6.12}$$

$$C_1 = \int_{\vartheta_{BE}}^{1} L(\theta) f(\theta) d\theta \tag{6.13}$$

$$C_2 = \int_1^{+\infty} L(\theta) f(\theta) d\theta \tag{6.14}$$

where the L(θ) is the loss or the benefit function and it has the following equation:

$$L(\vartheta) = \begin{cases} \mu a(1-\theta) - cr = \mu a(\theta_{BE} - \theta) & when \quad \theta < \theta_{BE} \\ 0 & when \quad \theta_{BE} \leq \theta \leq \theta_{BE} \\ cr - \mu a(1-\theta) = \mu a(\theta - \theta_{BE}) & when \quad \theta > \theta_{BE} \end{cases} \tag{6.15}$$

The expected loss or benefit is easy to compute by numerical integration instead of the calculation of the integral reported earlier.

A more conservative approach to the benefit cost analysis is to consider the neutral part as a benefit. In this way the distance between θ and θ_{BE} does not influence the difference between the benefit and the cost but it depends only from the value of the break-even point, and as it will be more extensive explained later, it depends on the cross site variance of the CMF.

6.2.1. *Comparison between the deterministic and the stochastic Benefit-Cost analysis for existing infrastructures*

To validate the methodology described above, and to know what the difference is on comparison with a deterministic approach it is applied on the before period of the CMFunction calibrated in the previous Chapter 5. The treatment is the retrofitting motorways with barrier meeting a new European standard.

For that segment, 68 in total, the Break – Even point is calculated using as cost for the treatment €200.000,00 per km and as cost of 1 ran-off-road crashes €400.000,00. The cost of the new barrier is taken by the construction cost list of ANAS S.p.A. [6.7], the owner of the most important Italian Network and the cost of crashes is taken by the "*Social crash cost estimation (2012)*" from the Ministry of Transportation [6.8] combining the percentage of fatal and injury crashes. The CMFunction is applied to each segment calculated on the value of CCR. As such, each segment has its own CMF and break - even point. The value of r chosen is 1, or better it is supposed that the Agency applies the treatment when the ratio between the benefit and the cost is equal or bigger than 1.

In this way for each segments the application of treatment is suggested if the CMF is smaller then the break – even point, on the contrary the treatment is not suggested. Appling the procedure described above for 38 segments the treatment is suggested while the other 30 segments that were treated, the treatment is not economically justified.

To validate the procedure the index D was used calculating using the Equation 6.7. Each segment has a different value of D, and the stochastic benefit-cost analysis as well as the deterministic one are applied on the segments which have the lowest and the higher value of D for the two condition: the treatment is suggested and the treatment is not suggested.

For calculating the cost and the benefit using the deterministic approach the difference between the benefit and the cost is evaluated, considering the benefit for the first year after the installation of the new barriers. This is why the application of the deterministic methods described earlier in the Chapter start from the initial cost and take into account the maintenance cost and the eventually other cost during the service life of the treatment. That cost are the same in the two approach, deterministic and stochastic. Moreover when a treatment has no maintenance cost, and the service life could not be easily evaluated, like the installation of barrier, the analysis on an indefinite service life could bring to an overestimation of the benefits. In the elaboration was not considered the service life for the treatment as well as the benefit. In the analysis the two different stochastic approach described earlier in the Chapter are applied to the 4 segments chosen for the elaboration. The results are reported for an ideal segment which has the characteristic of the segment considered for the elaboration but with a length of 1 km. As such it is possible to make a comparison of segments with different length. It was possible because in the computation of the expected number of crashes the length of the segment is considered as an offset and the overdispersion of the model is a function of the length.

In the Table 6.6 the value used in the analysis are reported for a standard segment of 1 km.

To evaluate the benefit and the cost for the stochastic approach the Equations from 6.12 to 6.14 are used. In the stochastic conservative approach the cost are considered starting from the expected value of the CMF. From an operative point of view, in this latter, the cost integral reported in the equation 6.13 has as lower limit exactly the expected value of the CMF.

Table 6.6. Value used for the benefit-cost analysis for the ran-off-road-crashes.

θ	μ	$θ_{BE}$	D	CCR	Std Deviation
0.3353	0.40	0.336	0.013	0.000	0.125
0.3353	0.39	0.3326	0.025	0.000	
0.3353	0.28	0.0577	2.605	0.000	
0.3333	6.00	0.9556	5.841	4.352	
r = 1					
c = € 200,000.00					
a = € 400,000.00					

As larger is the distance between the expected value of the CMF and the break even point the smaller is the probability to be wrong and it is reflected on difference between the benefit and the cost. Table 6.7 reports the results of the analysis for the 2 different approaches. The negative value are related to the higher value of the cost for the configuration of that segment.

Table 6.7. Comparison between the deterministic and Stochastic approach to the benefit-cost analysis for the ran-off-road crashes for the minimum and maximum value of D both where the treatment is suggested and where the treatment is not suggested (negative value).

D	Deterministic (B-C)	Stochastic (B-C)	Stochastic Conservative (B-C)
0,014	€ 436,66	€ 436.73	€ 436.73
0,025	-€ 807.47	-€ 807.40	-€ 807.40
2,606	-€ 58,911.89	-€ 37.063	-€ 58.912
5,841	€ 2,803,297.72	€ 1,778,496	€ 2,803,301

The difference in the results from the stochastic and deterministic approach take into account the perception of the benefit and the cost of either the Agency and the Community. In general the expected difference between the benefit and the cost is lower in the stochastic approach, to take into account that higher value than the expected value of the CMF is not really accepted by a users perception, while it represents still a benefit for the Agency.

The small difference between the deterministic approach and the stochastic conservative approach is related to the value of the variance and the distribution assumed for the CMF.

In the Gamma distribution, ended, the expected value is closer to the upper limit of the curves as much as the value of the variance is small. In this case the cross site variance of the CMF is relatively small and it is reflected on the shape of the curves and on the value of the B-C analysis. As the variance is small so the smaller the difference between the stochastic and the determinist approach.

In the Figure 6.2 the Equations from 6.12 to 6.14 are plotted for the first segment analyzed (D=0.014).

The B-C analysis is computed calculating the difference in the area described of the curves represented in Figure 6.2 respect to the x axis.

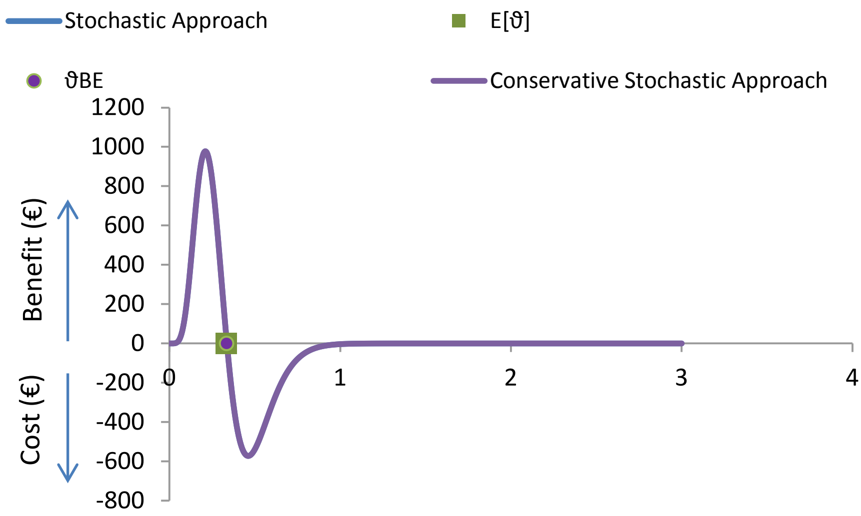


Figure 6.2. The benefit cost function multiplied to the probability function of the CMF.

6.2.3. Comparison between the deterministic and the stochastic Benefit-Cost analysis for new infrastructures

When a benefit-cost analysis has to be performed for a new infrastructure, or for an infrastructure in which the observed crash data are not available, the expected number of crashes cannot be evaluated.

In these cases both, the deterministic approach and the stochastic approach described earlier have not an high reliability because is not possible to apply the empirical Bayes correction and estimate the expected number of crashes on which the whole analysis is based.

However it is possible to think that the predicted number of crashes at a site, using an SPF calibrated on sites with similar characteristics, is a random variable with a mean and a variance.

Particularly the mean is the estimated value with the SPF $E(\mu_i)$, where i is the segment under investigation, and the variance is, considering a Negative Binomial error distribution (the expected value per site is Gamma distributed):

$$Var[E(\mu_i)] = E(\mu_i) + E^2(\mu_i) \cdot k \quad (6.16)$$

Where k is the overdispersion parameter computed for example using the equation from HSM [6.4] or from an SPF calibrated on site with similar characteristic when it is considered as variable with the length of segments (cfr. Chapter 5) [6.9] [6.10].

In this way, considering the same benefit or loss function described earlier in the Chapter (Equation 6.15) the benefit and cost are calculated with the following integral:

$$Cost \quad Function = \iint_{\mu,\theta} f(\theta) L(\mu;\theta) \cdot d\mu \cdot d\theta \quad (6.17)$$

In which L(μ;θ) has to be evaluated. The Equation 6.17 cannot be solved in the analytical way, but some approximations have to be introduced to solve numerically the integral. In the following an example with the proposed procedure is reported, using invented value for the computation.

To find a reliable solution of the equation 6.17 a Montecarlo simulation was performed using the SAS software package [6.11].

Monte Carlo methods (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results [6.12]; i.e., by running simulations many times over in order to calculate those same probabilities heuristically just like actually playing and recording results in a real casino situation: hence the name. They are often used in physical and mathematical problems and are most suited to be applied when it is impossible to obtain a closed-form expression or infeasible to apply a deterministic algorithm. Monte Carlo methods are mainly used in three distinct problems: optimization, numerical integration and generation of samples from a probability distribution.

Monte Carlo methods are especially useful for simulating systems with many coupled degrees of freedom, such as fluids, disordered materials, strongly coupled solids, and cellular structures. They are used to model phenomena with significant uncertainty in inputs, such as the calculation of risk in business. They are widely used in mathematics, for example to evaluate multidimensional definite integrals with complicated boundary conditions. When Monte Carlo simulations have been applied in space exploration and oil exploration, their predictions of failures, cost overruns and schedule overruns are routinely better than human intuition or alternative "soft" methods [6.13].

The modern version of the Monte Carlo method was invented in the late 1940s by Stanislaw Ulam, while he was working on nuclear weapon projects at the Los Alamos National Laboratory. It was named by Nicholas Metropolis, after the Monte Carlo Casino, where Ulam's uncle often gambled [6.14]. Immediately after Ulam's breakthrough, John von Neumann understood its importance and programmed the ENIAC computer to carry out Monte Carlo calculations.

In the elaboration the Montecarlo simulation was chosen because the difficulty in the computation of the Equation 4.35 considering both f(θ) and L(μ;θ) as random variable each with his own distribution. If a sample of f(θ) and L(μ;θ) is simulated n-time (thus n is equal to the number of simulation performed), a number n of differences between Benefit and Cost (B-C) is obtained. The average value of the n B-C value is be considered in this study the best estimator of the real B-C for that treatment and for that segment. The approximation is to consider the best value the average value and not the exact value of the interval chosen for the numeric computation of the integral. However if the number of simulation is large enough, that interval could be considered constant between a simulated value and the successive both for the CMF than for the predicted number of crashes.

In the example carried out below the value of benefit and cost for the first year after the implementation of the treatment are set as in the B-C analysis performed in the previous paragraph, although the treatment, and the segment as well are different.

Both the predicted value of crashes in the ideal segment and the variance are computed using the equations suggested by the HSM [6.4] for the rural multilane Highway in which the SPF for the base condition is:

$$E(\mu) = e^{a} * L * AADT^{b} \tag{6.18}$$

Where,

$E(\mu)$= base number of roadway segment accidents per year;

AADT = annual average daily traffic (vehicles/day) on roadway segment;

L = length of roadway segment (miles);

a, b = regression coefficients.

The value of the overdispersion parameter is determined as a function of segment length as:

$$k = \frac{1}{e^{(c+\ln(L))}} \tag{6.19}$$

Where,

k = overdispersion parameter associated with the roadway segment;

L = length of roadway segment (mi); and

c = a regression coefficient used to determine the overdispersion parameter.

The value of a, b and c are reported in HSM as well, as function of the severity level and are reported below in Table 6.8 .

Table 6.8. SPF Coefficients for Total and Fatal-and-Injury Crashes on Divided Roadway Segments from HSM.

Severity level	a	b	c
4-lane total	-9.025	1.049	1.549
4-lane fatal and injury	-8.837	0.958	1.687
4-lane fatal and injury (not poss. injury)	-8.505	0.874	1.74

For the elaboration a value of AADT of 50.000 vehicle per day is chosen on a length of 1 kilometer. Appling the Equations 6.18 and 6.19 to compute the predicted number of crashes and the overdispersion respectively, using the value of a, b and c from table 6.8 for 4-lane fatal and injury it is possible to calculate the variance of the predicted number of crashes using the 6.34. The results are shown below:

$$E(\mu) = \mathrm{e}^{-8.837} * 0.6 * 50{,}000^{0.958} = 2.76 \quad (6.20)$$

$$k = \frac{1}{e^{(1.687+\ln(0.6))}} = 0.308 \quad (6.21)$$

$$Var[E(\mu)] = 2.76 + 2.76^2 \cdot 0.308 = 5.12 \quad (6.22)$$

As it is clear from Equations 6.20 and 6.22 the prediction results to be overdispersed (mean lower than the variance). The index of effectiveness of the treatment considered in the analysis is 0.34. The expected reduction of crashes is 66% and the cross site variance of the CMF is equal to 0.12. Fixing the cost of the treatment equal to €200,000.00 per km and the crash cost equal to €400,000.00. The following expression was used for the deterministic B-C analysis, with the same meaning of symbols:

$$B - C = E\{\mu\} \cdot a \cdot (1 - E\{\theta\}) - c \cdot r = 2.76 \cdot 400{,}000 \cdot (1 - 0.34) - 200{,}000 \cdot 1 = €528{,}640.00 \quad (6.23)$$

The number of crashes calculated above are related to fatal and injury and to all the crash categories. The comprehensive cost of crashes has to be consistent with the categories of crashes expected for the segment under investigation.

The simulation was performed to evaluate the following expression:

$$B - C = [\mu_i \cdot a \cdot (\theta_j - \theta_{BE})] \cdot f(\theta_j) \cdot f(\mu_i) \cdot \Delta\theta \cdot \Delta\mu \quad (6.24)$$

where:

μ_i= is a random value of μ distribution;

θ_j= is a random value of θ distribution;

θ_{BE}= is the break-even point for the segment and the treatment;

$f(\theta_j)$= is the probability value related to the θ_j value, from the PDF of θ;

$f(\mu_i)$= is the probability value related to the μ_i value, from the PDF of μ;

$\Delta\theta$= average value of the step of variation of θ; and

$\Delta\mu$= average value of the step of variation of μ.

If the same approach used for the existing roadway segment is used the value of θ_{BE} has to be evaluated. Using the Equation 6.1 setting the value of r equal to 1 the break-even point for the treatment and the segment under investigation is 0,82.

The benefit-cost analysis was performed both with the stochastic approach and the conservative stochastic approach. Table 6.9 shown the results of the elaboration with 10^6 simulations.

Table 6.9. The results of the Benefit-Cost analysis for new roadway segments performed using a simulation of 10^6 value.

Approach	B-C Analysis (€/km)
Deterministic	€ 528,640.00
Stochastic conservative	€ 258,850.00
Stochastic	€ 192,874.00

As it is clear from the results on Table 6.9, the deterministic approach, which doesn't take into account the variability of the parameter used in the computation, give larger results than the both stochastic approach. Also in this case the distance between the break-even point and the expected value of the CMF is taken into account. Larger is the difference between the break-even point and the CMF lower is the value of the stochastic benefit cost analysis. Different conclusion arise when the stochastic conservative benefit - cost approach is analyzed. This latter is not influenced by the break-even point but it is strongly influenced by the variance of the CMF and the expected number of crashes. Larger are the variance and larger is the difference between the deterministic and the stochastic conservative approach.

6.3. Montecarlo simulation for the Benefit-Cost analysis of a combination of CMFs

The methodology could be extended using a combination of two CMFs. In this case the mathematical approach is too complicated to be solved using the traditional techniques. A Montecarlo simulation can help in the analysis. It is indeed possible to simulate the product of different CMFs. The distribution of the products could be obtained by a frequency analysis of the results, and the area can be calculated discretely summing for each frequency class the area under the curves. An approximation in that case is introduced, but it could be controlled by the class dimension in the frequency analysis. Smaller the class the higher the precision of the calculation of the probability that the product of more CMFs assumes the investigating values. With the aim to validate the methodology it was applied to the combination of two different CMFs.

The CMFs used for the analysis are one with index of effectiveness equal to θ=0.33 and cross site variance σ(θ) = 0.22, the other is the CMF studied by Hauer et al. [6.6] about the illumination for freeways θ=0.71 with the cross site variance σ(θ) = 0.06.

The frequency analysis results are shown below on comparison with a sample calculated ad hoc (Figure 6.3), in which a simulation with 5000 elements was performed for a single CMF.

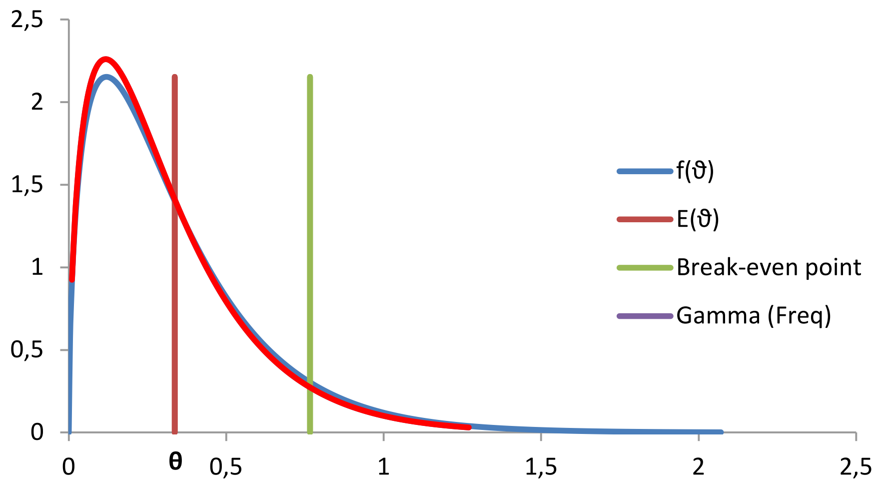


Figure 6.3. Comparison between the PDF of the f(θ) and the curves obtained by using the value of mean and variance from the simulation.

Using a class of 0.01 dimension the curves is calculated as a Gamma distribution with parameter α and β directly calculated from the simulated sample considering the same CMF using in the example (θ=0.33 and σ(θ) = 0.22) and a break-even point of θ_{BE} = 0.76.

The frequency analysis for a single CMF doesn't give good results. But it is not a big issue because the distribution of the CMFs is well known, and as shown in figure 6.3 the distribution calculated on the simulated sample gives adequate results. To corroborate this thesis Figure 6.3 shown the difference between the Normal and the Gamma distribution calculated using the sample parameter.

Another interesting result is delivered by the following figure (Figure 6.4) which shows a comparison between the curve related to the gamma distribution, in which the parameter of the distribution are calculated using the output sample of the simulation, and the normal distribution.

As shown in Figure 6.4 the normal distribution presents an overestimation of the benefit in comparison with the Gamma distribution. Considering the previous CMF, (θ=0,33 and σ(θ) = 0,22) and the same break-even point than the previous example (θ_{BE} = 0,76), the ratio of the benefit between the Gamma distribution and the Normal distribution is 0.75 (this latter indicates an overestimation of 15% of the Normal distribution). On the contrary the gamma distribution presents a loss overestimation of 10% more than a normal one. Thus, the normal distribution is not able to describe the product of two CMFs.

Despite the differences seen above the combination of more CMFs, such as the product of more Gamma distribution, can't be consider as a Gamma distribution as well and a frequency analysis is the only way to introduce the cross site variance in the reliability of the benefit-cost analysis. Unfortunately not all the studies presented in the literature, report the cross site variance of the CMF.

For the two CMFs introduced earlier in the paragraph the sample was simulated. 5000 simulations of the product were performed together with a frequency analysis.

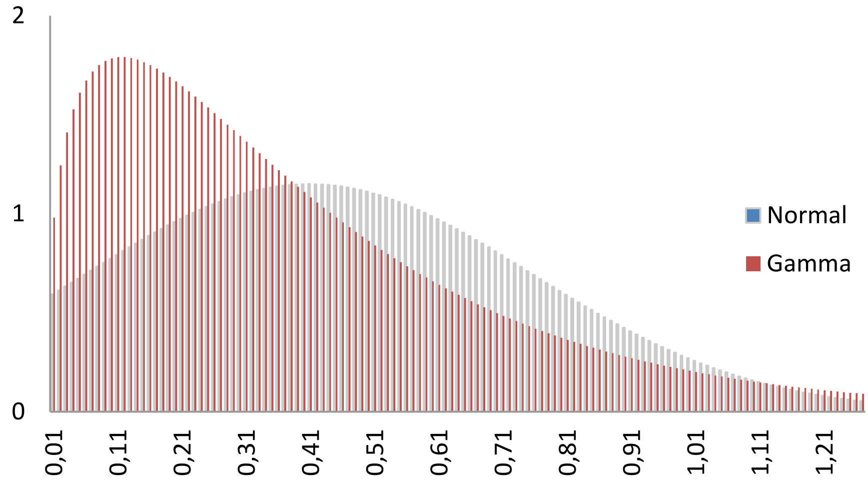


Figure 6.4. Comparison between the Gamma distribution obtained by using the values of mean and variance from the simulation and the frequency analysis of the data considering a normal distribution.

Calculating the expected value and the variance in the sample the following value are obtained:

$E(\theta)=0.13$ and

$\sigma(\theta) = 0.12$

in the frequency analysis the area of each element is calculated as follow:

$$\Delta A_j = \frac{C \cdot \sum_i N_{ij}}{n-1} \tag{6.24}$$

where:

- ΔA_j =area of the j class,
- C = class dimension,
- N_{ij}= number of element presents in the j class, and
- n= number of element present in the simulated dataset.

In figure 6.5 a graphic comparison between the frequency analysis and the Gamma distribution of the product of the two CMFs is reported.

In this case the frequency analysis shows that the product of the two CMFs is still comparable with a gamma distribution with the parameter α and β calculated from the sample. Checking time by time the distribution of the product of more CMFs or simply using the frequency analysis on the simulated dataset it is still possible to apply the procedure described earlier to calculate the reliability of the CMF for each site using the varying θ_{BE}.

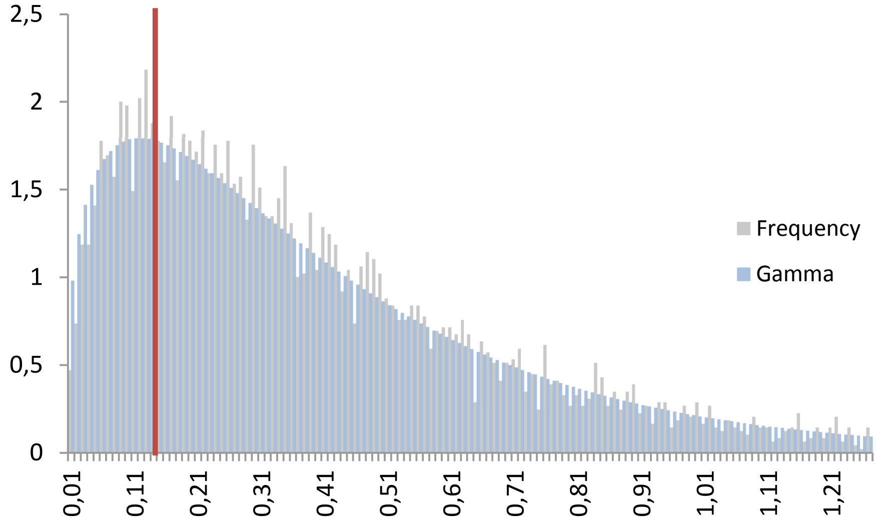


Figure 6.5. Comparison between the Gamma distribution and the frequency analysis on the simulated sample (the red line is the expected value of the product of the two CMFs).

6.4. Chapter summary

In the present Chapter 6 a second step of Hauer et al. research is reported "Value of Research on Safety Effects of Actions" [6.1]. Hauer et al. demonstrated that the reduction of the cross site variance could bring benefit reducing the possible loss due to the different results that applying a CMF in other contest, different than the site in which the CMF was calibrated. To do this a loss function was introduced which depends on the variance of the CMF. Taking a cue from Hauer et al research, extending the concept to possible benefit and the possible loss, a procedure was introduced either for existing road or new infrastructure. Particularly for new road a Montecarlo simulation was performed to add the uncertainty due to the predicted number of crashes, while for existing road the Empirical Bayes analysis was used. A comparison between the deterministic approach and a stochastic approach was performed.

References

[6.1]. Hauer, E., Bonneson, J. A., Council, F., Srinivasan, R., and Bahar, G.. Value of Research on Safety Effects of Actions. Transportation Research Record: Journal of the Transportation Research Board, No. 2280, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 68–74.

[6.2]. Hauer, E. Estimating the Mean and Variance of CMFs and How They Can Be Changed by New Research. Second report. Highway Safety Research Center, University of North Carolina, Chapel Hill, 2011.

[6.3]. Hauer. E.. Observational Before-After Studies in Road Safety. Pergamon Press. Elsevier Science Ltd.. Oxford. England. 1997.

[6.4]. Crash Modification Factor Clearinghouse. FHWA, U.S. Department of Transportation. http://www.cmfclearinghouse.org/. Accessed, September 25th, 2013.

[6.5]. NCHRP. Development of a Strategic National Highway Infrastructure Safety Research Agenda. In Task 1 report for NCHRP Project 17-48. Transportation Research Board of the National Academies, Washington, D.C., 2011.

[6.6]. Hauer, E., Bonneson, J. A., Council, F., Srinivasan, R., and Zegeer, C.. Crash Modification Factors, Foundational Issues. Transportation Research Record: Journal of the Transportation Research Board, No. 2279, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 67–74.

[6.7]. ANAS S.p.A.. Direzione Centrale Programmazione e Progettazione. Direzione centrale Lavori. Compartimento per la viabilità e la Sicilia, Palermo. Elenco Prezzi Unitari 2005.

[6.8]. Ministero delle Infrastrutture e dei Trasporti. Direzione Generale per la Sicurezza Stradale. Studio di valutazione dei Costi Sociali dell'incidentalità stradale. 2012.

[6.9]. Hauer. E.. Overdispersion in Modelling Accidents on Road Sections and in Empirical Bayes Estimation. Accident Analysis and Prevention. Vol. 33. No. 6. 2001. pp. 799–808.

[6.10]. Cafiso. S.. Di Silvestro. Persaud. B.. Begum. M.A.. Revisiting Variability of Dispersion Parameter of Safety Performance for Two-Lane Rural Roads. Transportation Research Record: Journal of the Transportation Research Board. No. 2148. Transportation Research Board of the National Academies. Washington. D.C.. 2010. pp. 38–46.

[6.11]. "SAS online documents", SAS Institute Inc. Version 8 February 2013.

[6.12]. WikiPediahttp://en.wikipedia.org/wiki/Monte_Carlo_method Access October 2013.

[6.13]. Anderson, H.L. (1986). "Metropolis, Monte Carlo and the MANIAC". Los Alamos Science 14: 96–108.

[6.14]. Baeurle, Stephan A. (2009). "Multiscale modeling of polymer materials using field-theoretic methodologies: A survey about recent developments". Journal of Mathematical Chemistry 46 (2): 363–426. doi:10.1007/s10910-008-9467-3.

[6.15]. Highway Safety Manual (HSM). First Edition. FHWA American Association of State Highway and Transportation Officials. U.S.A. 2010.

SUMMARY AND CONCLUSIONS

The presented work is the results of a PhD cycle deal with road safety for motorways (freeways in north America). Particularly data used in the elaboration are related to the motorway A18 Messina-Catania, in Italy. All the aspects related to the statistical approach to a safety management system were analyzed. The first step was the introducing of Safety Performance Function, which are the State of Art in the for the researcher for the identification of hazardous location, less, unfortunately for the Agencies in Europe. The problem is sometimes related to the lack of a common approach to the problem, covered in North America by the Highway Safety Manual, or lack of knowledge of the problem itself.

Different Safety Performance Functions were calibrated taking into account the time trend effects and others considering a priori lack of time correlation. The results, shown a strong effect of time trend when a long period of analysis is taken into account, and that the time trend effect has to be addressed when motorways are analyzed. The particularly roadway feature of motorways, the high speed and the lower crash rate, in comparison with other rural infrastructures amplify the time correlation effect which plays a fundamental role in the identification of hazardous sites. To do this the General Estimating Equation technique of calibration was used on comparison with the traditional Generalized Linear Modelling which don't consider the time trend effect in the estimation of model.

The second step of the present work was the optimization of the goodness of fit of the model varying the segmentation approach of the roadway segments. Five different segmentations were performed inserting different variables related to the road features and to the roadside hazard. The five segmentation tested were:

- All variable homogeneous (HSM approach);
- Two curves and two tangent inside each segment;
- AADT based;
- Curvature based; and
- Fixed length of segments (650 m was chosen because the maximum curve extension is 600 m, in this way not homogeneous segments are presents respect to the curvature).

The results of the estimation of the SPFs, calibrated using the Generalized Estimating Equation to account time trend, were tested using the Quasi-likelihood under independence Criterion (QIC) and with the CURE plot. In terms of goodness of fit the best results are given by the fixed length segmentation and by the segmentation which includes two curves and two tangent inside each segment. In this step of the research a ranking of the hazardous sites was performed using two different methodologies:

the Empirical Bayes correction and the Potential for Safety Improvement.

The top 10% hazardous location of the total segments was reported using the fixed length segmentation. To address the problem of the lack of the overdispersion parameter given by the methodology of calibration of the SPF needed for the ranking analysis a Generalized Linear Modelling approach was used to obtain the dispersion of data. Indeed the Generalized Estimating Equation, is a quasi-likelihood methodology and the estimation doesn't provide the dispersion of data. The common approach of the great part of the Agencies in Italy is to use the observed crashes to rank the managed road network. The performed ranking give different results than the observed and even some difference were found between the EB and the PSI methods, enforcing that the regression to the mean effects are always presents where crash events are analyzed and that the simple observed number of crashes in a certain period is not a reliable indicator of the real safety performance of a site.

The third step of the present study deals with the Crash Modification Factors. A CMF for a class of barrier meeting the new EU standard was reported, investigating the safety benefit on the ran-off-road crashes. The roadside safety is one of the main issue in Europe above all when motorways, high speed infrastructure, are analyzed. The methodology used was the before after empirical Bayes analysis, able to address the regression to the mean effect on an ad hoc segmentation, useful to isolating the variable under investigation. The segments taken into account were those where the new barrier was installed for the lateral, median or both.

A cluster analysis was introduced grouping the segments based on their Curvature change rate. Two different regression analysis, the linear and the exponential, were performed on the clustering to study the relationship between the CCR and the reduction of crashes due to the installation of the new class of barrier, evaluating the regression performance with R^2 coefficient of determination.

The exponential regression technique give the best results in the analysis with the higher value of R^2 and assessing a strong relationship between the curvature and the ran-off-road crashes. The Crash Modification function obtained could be used for the ran-off-road crashes and for fatal and injury severity class. For the total crashes the regression doesn't give reliable results with a value of R^2 lower than 0.5.

The last step of the analysis deals with the benefit- cost analysis. As consequence of the previous research in present research work, the evaluation of alternatives have to be assess as conclusion of a safety analysis and evaluation of alternatives. The approach used is a possible extension of a methodology introduced by Hauer et al. in two different studies published on Transportation Research Record in 2012. The first of those, titled "Crash Modification Factors, Foundational Issues" reported a methodology to investigate the transferability of CMFs when more than one study exist in literature on the same treatment, and introduce a new concept of variance. Hauer et al, argued logically, that not only the variance of the CMF has its importance, but also the variance due to the application of the CMF in different site. The combination of the two variance is named “cross site variance” in the present work. Using the concept of cross site variance, the variation of the results of CMF calibration on the cluster analysis was used to calculate the additional term for the variance.

In the second study titled “Value of Research on Safety Effects of Actions” Hauer et al. demonstrated that the reduction of the cross site variance could bring benefit reducing the possible loss due to the different results that applying a CMF in other contest, different than the site in which the CMF was calibrated. To do this a loss function was introduced which depends on the variance of the CMF. Taking a cue from Hauer et al research, extending the concept to possible benefit and the possible loss, a procedure was introduced either for existing road or new infrastructure. Particularly for new road a Montecarlo simulation was performed to add the uncertainty due to the predicted number of crashes, while for existing road the Empirical Bayes analysis was used. A comparison between the deterministic approach and a stochastic approach was performed.

The results shown a general overestimation of the benefit of the deterministic approach, enforcing the idea that considering the variance even in the benefit-cost analysis could bring a more reliable results on the evaluation of the different alternatives and can play a fundamental role in the decision making process for the Agencies.

Highlights and future work

The present research work tried to address the problem related to the safety analysis for roadway segments focusing on the reliability of the benefit-cost analysis for the evaluation of the effects of a treatment. The Highlights of the work are listed below:

- All the variables considered in a traditional benefit-cost analysis have a stochastic nature but they are considered in a deterministic way;
- The combination of the cross site variance of the CMFs and the variance estimated using the empirical Bayes methodology, assesses the variability of the CMF in site with different characteristics, and allows a statistical inference on the Gamma distribution of the CMF itself;
- A stochastic approach to the benefit cost-analysis shown a general overestimation of either the benefit or the cost then the deterministic approach;
- The introduction of the break-even point gives to the Agency the possibility of introduce an important index to evaluate the reliability of the treatment;

- The stochastic benefit-cost analysis has different characteristic when it is applied to new or existing infrastructures;
- When one is dealing with new infrastructure in the evaluation of the benefit and cost the target crashes related to the treatment have to be considered as a random variable as well because the low reliability in their computation;
- The combination of more CMFs can be considered as well in a stochastic benefit-cost analysis introducing a Montecarlo simulation for the product and computing the results with a frequency analysis.

Although encouraging results the proposed research work needs to be tested on existing project to validate the results and to improve the methodology. Furthermore the difficulty in the analytical computation of the benefit and cost may be overcome introducing an algorithm able to solve the equation automatically using calculators. The combination of two CMFs could be considered in the stochastic benefit-cost analysis but it is outside of the proposed methodology the test of the combination of more than 2 CMFs.

Finally the procedure presented in Chapter 5 to evaluate the cross site variance of the estimated CMF has a limitation in this application. The cross site variance has to be able to describe how the CMF varying from a site to another. In other terms it is able to catch the different conditions (related e.g. with road feature, traffic composition, geometry, etc.) present in different segments. When it is applied on the cluster analysis the value of the cross site variance could be underestimated or overestimated, because the variability related to one of the more important variable, which influence the target crashes, the curvature in that case, is taken into account.

Furthermore a great variability form a site to another when roadway segments are analyzed is given by the different lengths of segments. To eliminate the influence of the length of segments a fixed length segmentation approach has to be tested.

Finally the variance of Crash Modification Function need to be evaluated if it is used in the benefit-cost analysis. In the proposed research work a methodology was developed on the cluster analysis, but it underestimates the value of the cross site variance and more studies need on the topic.

A further analysis could be conducted is to evaluate how the cross site variance of the CMF can influence the transferability and how it can be used to improve the evaluation of the effects of a treatment in different site. Further studies needed to assess if under certain circumstances the cross site variance can be considered as a reliability index for the transferability of the CMF.

Although, more research needed to address the problem of the evaluation of alternatives using a stochastic benefit-cost analysis, the present study may represent a starting point for the topic and a motivation for future work.