H. Hautzinger
Institute of Applied Transport and Tourism Research (IVT), Heilbronn, Germany

Multi-Level Statistical Models for Vehicle Crashworthiness Assessment – An Overview

Abstract

Empirical vehicle crashworthiness studies are usually based on national or in-depth traffic accident surveys: Data on accident-involved cars/drivers are analysed in order to quantify the chance of driver injury and to assess certain risk factors like car make and model. As the cars/drivers involved in the same accident form a 'cluster', where the size of the cluster equals the number of accident-involved parties, traffic accident survey data are typical multi-level data with accidents as first-level or primary and cars/drivers as second-level or secondary units (car occupants in general are to be considered as third level units). Consequently, appropriate statistical multi-level models are to be used for driver injury risk estimation purposes as these models properly account for the cluster structure of traffic accident survey data. In recent years various types of regression models for clustered data have been developed in the statistical sciences. This paper presents multi-level statistical models, which are generally applicable for vehicle crashworthiness assessment in the sense that data on single and multiple car crashes can be analysed simultaneously. As a special case of multi-level modelling driver injury risk estimation based on paired-by-collision car/driver data is considered. It is demonstrated that assessment results may be seriously biased, if the cluster structure inherent in traffic accident survey data is erroneously ignored in the data analysis stage.

Introduction

Vehicle crashworthiness, i.e. the ability of a vehicle to protect its own occupants in collisions, is an important research subject in the traffic safety sciences. It is, of course, a theoretical concept which can be specified in several different ways. Basically, vehicle crashworthiness is to be measured by occupant injury risk given accident involvement. However, as injury information may not necessarily be available for all occupants of accident-involved vehicles, crashworthiness is frequently quantified simply by driver injury risk: the higher the driver's probability to be injured, the lower the vehicle's crashworthiness.

The injury status of accident-involved drivers is affected by many characteristics of the driver and his or her vehicle. In addition, factors like collision speed, mass of opponent vehicle and so forth play an important role. Among the numerous determinants of driver injury risk, the risk factor 'make and model of the car driven' is of special and sometimes even primary interest. The latter will, for instance, be the case in car safety rating based on real-world crash data. Although this paper focuses on the assessment of car make and model as a determinant of driver injury, the methods presented here can easily be transferred to the analysis of any other risk factor of interest.

Investigations in vehicle crashworthiness may be viewed as special cases of epidemiological studies as they deal with the distribution and determinants of a specific 'disease' (=injury due to accident involvement) in a specific human population (=accident-involved car drivers). Descriptive crashworthiness analyses are conducted in order to estimate the average safety level of different car models; the results of such analyses may, for instance, be of interest for the motor car insurance industry. Analytical (also termed 'aetiological') studies designed to measure the partial effect of car make and model on driver injury risk preferably correspond to the car buyer's perspective.

Partially based on results of the SARAC project [1] the following topics are treated in the sequel:

- Estimating absolute and comparative chance of driver injury for cars grouped by make and model.
- Testing association between risk factor car make and model and criterion variable car driver injury status (crashworthiness comparisons between different groups of cars).
- Measuring comparative chance of car driver injury in the case of multi-level data (random effects probit and fixed effects logit models).
- Adjustment of group-specific injury risk rates for third variables (confounding factors).

- Paired-by-collision car/driver data as a special case of multi-level traffic accident data.

The paper gives an overview of elementary and more advanced statistical methods for vehicle crashworthiness assessment taking explicitely account of the multi-level structure of traffic accident data. The assessment of vehicle aggressivity defined as the degree to which injury is inflicted upon occupants of the vehicle or road user with which the 'subject' car crashes will not be treated here.

Basic Concepts

Accident survey as the basic study type for crashworthiness investigations

Empirical crashworthiness investigations belong to the class of observational studies: data are collected on real-world accidents and accident-involved cars which provide information on car make and model and car driver injury status but also on third variables which might affect driver injury risk. Typically, a certain traffic accident survey[1] is the data source and thus the study design of a crashworthiness investigation may be characterized as an ex-post-facto approach.

In many practical situations these surveys will have been conducted by a national statistical office and thus the accident survey under consideration will frequently be a complete census in the sense that all police-recorded accidents are contained in the database. In some cases, however, the empirical data may have the character of a sample, e.g. when they have been collected in a local or regional in-depth accident investigation under a specific sampling plan. As usually for each accident recorded in the survey empirical data on all accident-involved cars are collected, a traffic accident survey normally is a cluster survey with accidents as primary and accident-involved cars as secondary units.

Surveys are referred to as cross-sectional investigations when they provide data for a limited study period. In the case of traffic accident surveys, however, the system of data collection can frequently be considered as a permanent survey. Therefore, incidence[2] of injury (number of cases of injury within a specified period of time) caused by traffic accidents may be measured for consecutive time periods (e.g. weeks, months or years).

Measuring the risk factor car make and model

In epidemiology[3] any potential determinant of the disease under study is termed risk factor. In the context of crashworthiness assessment, of course, car make and model is the risk factor of primary interest. It is important to note that 'car make and model' is a theoretical construct which needs to be specified carefully. Obviously, one may consider car make and model as a complex attribute of a vehicle summarising all its physical and design properties. In this case vehicle characteristics like length, mass and body style of car are already 'contained' in the vehicle attribute car make and model. When the above concept is applied, it would make no sense to consider crashworthiness as a function of car make and model and mass of car, as car mass is simply one of the constituent properties of a given car model.[4] Under such a perspective the car mass effect on crashworthiness (mass of the 'subject' or 'focus' car itself, not mass of the opponent car!) simply cannot be separated from the effect of car make and model.

Alternatively, car make and model may be considered as an attribute characterising only the car's design properties and secondary safety fittings. In this case the car's purely physical properties like length and mass could be assumed

1 In addition to surveys, so-called case-control studies are also widely used retrospective observational types of investigation. As case-control study is always based on two different samples, one sample of 'cases" (persons with a specific disease) and a second sample of 'controls" (persons without the disease). These two groups are compared with regard to the risk factor. From a methodological point of view surveys are normally preferred to case-control studies, especially when the survey is a representative sample from the population of interest or even a complete census. It will, however, be shown later in this paper that it might well be advantageous to build a database from a standard traffic accident survey which formally corresponds to the database of a so-called matched case-control study.

2 In epidemiological studies, in addition to incidence one is often also interested in the prevalence of a certain disease (number of existing cases of disease at a particular point in time). As accidents are events in time and space and not objects or subjects (like patients), the concept of prevalence does not apply to crashworthiness investigations.

3 The basic concepts and methods of epidemiology are presented here following [2].

4 For instance, when drugs A, B,... are compared in clinical trials it is not common to consider the amount of a certain active substance contained in the drugs as a separate variable. Consequently, no adjustments or corrections for this variable are made.

to be separate determinants of vehicle safety. Such a concept might be useful if car make and model is defined in the broader sense of 'car model group' such that cars belonging to the same model group may differ in mass (variation of vehicle mass within car model group). However, even if we assume car mass and car model group to represent different aspects of a vehicle, it is clear that car mass is largely determined by car make and model. As this dependency may be considered as a causal relationship ('the vehicle is heavy because it is a Mercedes S class'), it is at least questionable whether driver injury risk should be adjusted for mass effects in order to measure the pure effect of the car's structure, design and secondary safety fittings on occupant protection.

Obviously, the risk factor car make and model is a categorical variable. The main purpose of vehicle crashworthiness assessment is to estimate an appropriate index of crashworthiness for each category of cars. In the simplest case only two categories are distinguished. In this situation one often speaks of 'subject' car model (e.g. VW Golf 4) and 'other' car model (e.g. not VW Golf 4) corresponding to the usual 'exposed/unexposed' dichotomy. Sometimes, however, we may explicitely consider several different car models which means that the risk factor under consideration has a whole set of possible categorical (i.e. unordered) outcomes. It is not uncommon to distinguish up to about 150 different car models in a single crashworthiness study.

Measuring car occupant injury status

Vehicle crashworthiness has been defined as the ability of a car to avoid injury to its own occupants in collisions. Thus, occupant injury status is the dependent or criterion variable of any crashworthiness study. 'Injury', of course, may be measured in quite different ways ranging from coarse classifications based on police reports to rather complex injury scales (AIS, ISS). Ideally, one would like to have a clinical definition of what is meant by 'injury' which can be tested by objective evidence. In many practical situations, however, the vehicle safety analyst has to rely on police reports or insurance files, where the validity and consistency of diagnostic criteria is at least questionable.

One could, of course, decide to use mortality (death due to crash involvement) as the criterion variable. This, however, would result in a significant reduction in the number of cases of 'injury' to be found in the population of accident-involved car occupants. Therefore, one usually moves further down the hierarchy of severity of injury:

- died,
- hospitalized,
- diagnosed or self-reported accidental injury.

Statistical analysis of driver injury risk is considerably simplified if the crude binary attribute 'driver injured: yes/no' is used to describe accident outcome. One can, however, also distinguish several (ordered) levels of injury severity ranging from 'uninjured' to 'killed' in the case of police-recorded data or from AIS 0 to AIS 6 when in-depth data are available.

It should be noted that crashworthiness assessment results may depend on the accident outcome measure adopted. Therefore, careful consideration of the injury status variable used is necessary when comparing different vehicle safety assessment approaches. As already stated above, vehicle crashworthiness is frequently measured simply by car driver injury. Only in a few studies both driver and (front) passenger injury is considered. Not surprisingly, the choice of the accident outcome variable is often dictated by data availability.

Measuring the chance of driver injury incidence

Vehicle crashworthiness assessment aims at evaluating the chance that an accident-involved driver of a certain car model (the 'subject' car model) is injured. The most basic epidemiological measure is the risk of injury, i.e. the probability of an accident-involved driver being injured given that he or she drives a 'subject' car. Like in other fields of safety research, however, various alternative risk concepts are applicable in vehicle crashworthiness studies (absolute and relative risk, odds and odds ratio).

Classical methods for estimating risk measures from survey data are based on the assumption that the study units (in our case accident-involved vehicles/drivers) have been selected by simple random sampling. In the following chapter it will be demonstrated that this assumption is by no means valid in crashworthiness studies. Rather, the study units are grouped or clustered in a natural way. This characteristic property of empirical traffic accident data calls for specific methods of risk estimation.

Crashworthiness Assessment Based on Multi-Level Car/Driver Data from Traffic Acident Surveys – The General Case

Multi-level structure of the population at risk

In studies on vehicle crashworthiness the universe of accident-involved cars or, more precisely, the universe of ‘accident involvements of cars’ has to be considered as the population at risk. Of course, this population must as usual be well defined with respect to factual, spatial, and temporal characteristics. It is important to note that the elements of the population at risk are neither fixed subjects nor objects but rather events occurring in time and space.

Obviously, a collision of two cars corresponds to two different accident involvements of cars which, however, refer to the same accident. In this situation one may speak of ‘paired-by-collision vehicle/driver data’. More generally, it can be said that the population at risk has a multi-level structure with accidents (single and multiple vehicle accidents) as first-level units, accident-involved vehicles as second-level units and occupants of accident-involved vehicles as third-level units. The first level units, i.e. the accidents, may be considered as clusters of accident-involved cars where the size of the cluster (1, 2, 3,…) corresponds to the number of parties involved in the crash. This multi-level or cluster structure of the population at risk must be taken into account in any methodologically sound crashworthiness investigation.

As a prerequisite for vehicle crashworthiness assessment, information on the following characteristics is needed for each selected element of the population at risk, i.e. for each accident-involved car in the sample:

- injury status of car driver (criterion variable),
- make and model of car (risk factor to be assessed),
- other factors which might affect the criterion variable (concomitant variables).

Risk and relative risk of car driver injury

The purpose of crashworthiness assessment is to evaluate the chance of car driver injury in case of an accident. An appropriate quantitative measure is the risk of car driver injury, i.e. the probability of the car driver being injured given that the car belongs to a particular group of vehicles (‘subject’ car model or, more generally, group of ‘exposed’ study units). The risk of car driver injury (also termed absolute risk) describes the relationship between a specific car make and model and car driver injury status. This, however, is not sufficient for assessing the risk factor to injury outcome.

As in other fields of evaluation research a comparison group is required which, for instance, may be the group of accident-involved cars which do not belong to the car model category under consideration (‘other’ car model or, more generally, group of ‘unexposed’ study units). This leads to the definition of relative risk as the ratio of injury risk for the drivers of a particular car model to the injury risk for the drivers belonging to the comparison group. When several different car models are considered simultaneously, one category is chosen to be the base or reference category and the analyst compares all other categories to this base.

Odds and odds ratio of car driver injury

However, risk as a probability is not the only possibility of specifying ‘chance’. An alternative specification is called the odds. The odds measure the number of times accidental injury occurs relative to the number of times it does not. The odds can be calculated for different groups of vehicle/driver units. In car safety rating one is interested in the ratio of the odds for a particular car model to the odds of a comparison group which, for instance, may consist of all other car models. The corresponding measure is called odds ratio.

It is important to note that in epidemiology and other fields of applied statistics researchers take ‘odds’ and ‘odds ratio’ to refer to the chance of disease (accidental injury) incidence just as they do ‘risk’ and ‘relative risk’. In practice, the odds as such are rarely of interest, and the odds ratio is generally quoted alone. In vehicle crashworthiness studies one can, of course, think of several alternative ways to measure car driver injury risk. Clearly, however, only statistically sound concepts of specifying ‘chance’ of occupant injury are acceptable. Careful interpretation of the risk measure used is necessary.

Relative risk and odds ratio as measures of association between car make and model and car driver injury

It should be stressed again that the relative risk and the odds ratio are meaningful measures of association between risk factor car make and model and car driver injury. Both quantities measure the relative chance of driver injury for a particular subject car model, compared to a certain base category or reference group of cars: if the crashworthiness of the subject car model does not differ from the crashworthiness of the reference group of cars, one can expext the relative risk and the odds ratio to be around unity (corresponding to 'no association' between risk factor and driver injury status).

As crashworthiness assessment normally aims at some ranking of car models, it can be said that the calculation of the relative risk and the odds ratio is fundamental to any car safety rating system. Obviously, both the relative risk and the odds ratio have to be estimated from empirical data collected in accident surveys. Subsequently, it will be shown that due to the specific nature of the population at risk (clustering of study units) the classical approaches to measuring relative chance of driver injury are not suitable.

Estimating relative risks and odds ratios in the case of multi-level data

Why the classical (unmatched) approach is not appropriate

According to epidemiological standards a given categorical risk factor is assessed by computing a 95% confidence interval for the population value of the relative risk or the odds ratio (subject car model compared to reference car model). If the confidence interval does contain unity, one concludes that the risk factor car model has no effect on driver injury. If the lower (upper) limit of the confidence interval is above (below) unity, one concludes that the injury risk for subject car model drivers is higher (lower) compared to the reference group of cars. For computational details see [2], Chapter 3.

The classical approach outlined above assumes that a simple random sample of units has been drawn from the population at risk. Under simple random sampling the sample inclusion probability of a specific unit a is not affected by the drawing of some other unit b from the population. This, however, is not the case in vehicle crashworthiness studies due to the multi-level structure of the population. If, for instance, one car involved in a specific two-car accident is in the sample, the second car involved in the same accident will automatically be also in the sample. Thus, for any two-car crash the corresponding accident involvements of cars (say car 1 and car 2) must be considered as a 'cluster' in the sense that – irrespective of car make and model – the injury status of the two drivers is not independent: if driver 1 is injured (uninjured), driver 2 also tends to be injured (uninjured) for obvious reasons. The same argument holds for accidents where more than two vehicles (more generally: 'parties') were involved.

Risk factor assessment based on multi-level models

As there is a known clustering within the accident involvement data, a multi-level model will be appropriate for assessing the risk factor car make and model. In the context of crashworthiness assessment multi-level modelling means that regression models with accident-specific parameters are used. These accident-specific parameters which sometimes are also called 'effects' may be assumed to be fixed or random.

If the driver injury status variable is binary, the fixed effects logit model or the random effects probit model can be applied with driver injury status as the dependent and risk factor status (car make and model) as the explanatory variable. See, for instance, [3] and [4]. As usual, the different categories of car make and model are represented by dummy variables in the logit or probit model. From these models one obtains point and interval estimates of the relative chance of driver injury for one or more 'subject' car models compared to a given 'reference' car model. In contrast to the classical approach these estimates account for the clustering within the accident involvement data. Another substantial advantage of these models is that additional variables can be included thus adjusting risk estimates for confounding factors. Subsequently, the random effects probit model is briefly described.

Let accidents be labelled by the index $i(i=1,\ldots,n)$ and the cars within accident i by the index where $j(j=1, \ldots, m_i)$ where $m_i=1, 2, 3_{,\ldots}$. Then, for the observable dichotomous driver injury status variable, i.e. for the dependent variable

$$y_{ij} = \begin{cases} 1 & \textit{if driver } j \textit{ in accident } i \textit{ is injured} \\ 0 & \textit{otherwise} \end{cases}$$

a probit model with random effects could be developed, the structure of which is

$$y_{ij} = \begin{cases} 1 & \textit{if } y_{ij}^{*} > 0 \\ 0 & \textit{otherwise} \end{cases}$$

where

$$y_{ij}^{*} = \mu + \sum_{k} \beta_k x_{ijk} + u_{ij}$$

is a non-observable ('latent') continuos variable to be interpreted, for instance, as a combined index of strength and effect of the forces acting upon driver *j* in accident *i*. If this index exceeds a certain threshold value (zero), the event 'driver is injured' will be observed. The latent variable is assumed to depend on a set of indicator variables χ_{ijk} (k=1, 2, …) attaining the value 1 or 0 if category *k* of the risk factor car make and model is present or not present at the *j*-th car within the *i*-th accident and a random error term

$u_{ij}=a_i + \varepsilon_{ij}$

consisting of an accident-specific random component a_i and a purely random component ε_{ij}. As can be seen, a_i allows for random variation of the latent variable at the accident level, whereas ε_{ij} accounts for random effects at the car/driver level. Regarding the random effect a_i it is assumed that this variable is normally distributed with mean 0 and variance σ_{α}^{2}. As usual it is assumed that the component ε_{ij} is normally distributed with mean 0 and variance $\sigma_{\varepsilon}^{2} = 1$.

The above variance component model implies that for any given accident *i* the latent variables γ_{ij}^{*} (*j*=1,... m_i) and thus the observed injury outcomes of the drivers involved in the same accident are positively correlated with correlation coefficient $\gamma = \sigma_{\alpha}^{2} / (\sigma_{\alpha}^{2} + 1)$. This property makes the model suitable for analysing clustered data as is the case in crashworthiness assessment.

Using empirical crash involvement data (*n* accidents, $m = \sum_{i=1}^{n} m_i$ accident-involved cars/drivers) one can estimate by the maximum likelihood method

- the model constant μ
- the parameters β_k associated with the different categories of the risk factor car make and model and
- the coefficient of correlation γ.

By definition, the parameter associated with the reference car model category is equal to zero. Therefore, a positive (negative) sign of β_k indicates that the accident consequences are more (less) severe for drivers of car model category k compared to drivers of the reference group of cars. As the standard errors of the parameter estimates can be estimated, it is possible to compute confidence intervals and to test hypotheses about the parameters and thus hypotheses about the crashworthiness of different car models.

Models of the type outlined above can be estimated using appropriate statistical software. The author has estimated random effects probit models by means of LIMDEP procedure PROBIT in a mobility behaviour study (car use of persons within households as a function of characteristics of the person and the household). See [5]. The problem structure in crashworthiness studies is quite similar (injury of drivers within accidents as a function of characteristics of the car/driver and the accident).

It appears that the estimation of the relative chance of driver injury may lead to erroneous results if the clustering within the accident involvement data is not taken into account in the data analysis stage. This will be shown in the following chapter, where a practical example is presented. To the author's knowledge, general multi-level models of the type described above have not yet been applied in vehicle crashworthiness assessment using at the same time data on single car, two-car and multiple car accidents. However, for the important special case of two-car accident data ('paired-by-collision vehicle/driver data') examples of multi-level models can already be found in the literature which will be quoted later.

Crashworthiness Assessment Based on Paired-by-Collision Car/Driver Data

Rationale for using two-car accident data only

Several existing vehicle safety rating methods[5] restrict themselves to analysing data on two-car accidents only. From a methodological point of view this approach may be interpreted as a so-called 'matched pairs design' frequently encountered in epidemiological studies. Typically, the restriction to two-car accidents is to minimise distortion which would be caused to the estimates of driver injury risk if, for instance, a particular car model had a high proportion of collisions with much larger vehicles such as lorries or busses. As will be seen below there are, however, even more convincing arguments in favour of the use of paired-by-collision car/driver data for crashworthiness assessment purposes.

As already stated above the restriction to two-car accidents corresponds to a matched pairs design (also termed '1:1 matching' or 'pair matching'): the cars involved in the same accident are considered as a single matched pair rather than two independent observations. Considering pairs of cars has the advantage of high internal validity[6] since all observed and unobserved (!) characteristics of the accident itself (time, location, weather conditions etc.) are the same for both accident-involved cars and, therefore, these characteristics cannot account for possible differences in the injury risk of the two drivers involved in the accident. Consequently, 'confounding' is reduced and the 'pure' effect of car make and model on the chance of car driver injury can be measured more precisely.

Another very attractive feature of the matched pairs design is that it can equally been applied to accident databases with and without damage-only accidents. This is because accidents where both drivers are not injured or both drivers are injured (so-called concordant pairs) tell nothing about the relative risk of driver injury (subject car model compared to other car model). Practically all in-depth accident databases and many police-recorded accident data sets do not cover accidents with material damage only. In all these cases it is obviously not possible to estimate the absolute risk of driver injury because a substantial part of the accident-involved non-injured drivers is missing in the accident database. All one can do is to estimate the comparative chance of driver injury which can best be accomplished under a matched pairs design.

When adjustment for confounding is made at the design stage of the study by choosing the concept of matched pairs, this must be taken into account in the stage of data analysis. A matched pairs study requires a matched pairs analysis, which can be more complex both to understand and compute. See [2], Chapter 6, for a general presentation. For statistical details of matched studies in vehicle safety research see [1]. As an example of proper application of matched pairs analysis methods (conditional logistic regression models for paired data) in the context of estimating the comparative chance of driver injury see [7].

The decision on the statistical method to be used for data analysis depends on the answers to the following questions:

- Is the assessment of the risk factor under consideration to be made without or with adjustment for confounding car- and driver-specific[7] variables?
- Is driver injury status a binary variable or is it measured on an ordinal scale with several levels?
- Are we mainly interested in testing association between risk factor and driver injury?
- Is estimation of the comparative chance of driver injury a main concern of the study?

Depending on study purposes and scaling of the variables, the statistical tool box offers various methods for vehicle crashworthiness assessment in the case of paired-by-collision car/driver data.

The candidate approaches can be broadly classified into statistical models with population-averaged and models with accident-specific parameters. The two approaches differ in the way of modelling the dependence between the injury status of the two-car drivers belonging to the same accident. The first approach leads to the class of log-linear models of driver injury in two-car

---

5 A detailed description of existing car safety rating methods can be found in a specific SARAC report prepared by the author. See [6]. It can be said that all safety rating methods described in this report ignore the clustering of car/driver data.

6 Obviously, the external validity of vehicle crashworthiness assessment is reduced by analysing paired-by-collision car/driver data only, since a substantial part of the population of all accident involvements of cars (especially single car crashes, crashes against freight transport vehicles etc.) is ruled out. As the severity of these crashes tends to be above average, this is an obvious weakness of the matched pairs design.

7 Adjustment for accident-specific covariates is automatically made due to matching.

accidents, the second preferably to so-called fixed effects models of driver injury for cars/drivers matched in pairs.

In various fields of applied statistics fixed effects models proved to be well suited for analysing matched pairs data, especially data from matched case-control or matched cohort studies. It turnes out that this is also valid for crashworthiness studies based on two-car accident data.

Crashworthiness assessment without adjustment for car- and driver-specific variables

When no adjustment for confounders is to be made, paired car/driver data can always be displayed in a square two-dimensional contingency table, where the two dimensions of the table (rows and columns, respectively) correspond to the two accident-involved cars/drivers. If driver injury status as well as car make and model are binary variables, a 2x2 table will arise. For the construction of this 2x2 table two different possibilities of crosstabulating accidents exist:

- Matched cohort study design: Accidents by injury status of subject car driver (rows) and injury status of other car driver (columns).
- Matched case-control study design: Accidents by car make and model of injured driver (rows) and car make and model of uninjured driver (columns).

The null hypothesis of no association between the risk factor car make and model and the criterion variable car driver injury status can be tested using a so-called symmetry test. When dealing with 2x2 tables McNemar's test will normally be appropriate.

When more than two levels of driver injury are to be distinguished, the matched cohort design is appropriate. If more than two categories of car models are to be assessed, the matched case-control design with injured drivers as cases and uninjured drivers as controls is to be chosen by the analyst. In both situations the empirical accident frequency data can be displayed in rxr tables. For rxr tables Bouwker's test is the appropriate statistical method for testing the hypothesis of no association.

For matched studies the odds ratio is the generally accepted measure of comparative chance of driver injury (injury odds for subject car driver divided by injury odds for other car driver). Under a matched pairs design the odds ratio can be estimated from the corresponding 2x2 table. It turns out that for paired data the estimate of the odds ratio[8] only depends on the two off-diagonal elements of the 2x2 table[9]. The corresponding estimate is often termed matched odds ratio. When under a matched case-control study design several car models are to be distinguished, the empirical accident frequency data are displayed in a rxr table. In this situation the odds ratio can be calculated for all combinations of car models (e.g. combinations A/B, A/C and B/C when three car models A, B, C are considered).

Under a matched design confidence intervals for the population value of the odds ratio can be computed using the F-distribution. It is important to note, that the choice of the odds ratio estimate (matched odds ratio versus cross-product ratio) does not only affect the confidence interval but also the point estimate of the driver injury odds ratio. When pairing is ignored, the odds ratio estimate may be seriously biased. See the example below.

Crashworthiness assessment with adjustment for car- and driver-specific variables

When the adjusted odds ratio for the risk factor car make and model is of interest (adjustment for confounding car- and driver-specific variables), the above analysis of two-dimensional contingency tables is no longer appropriate. Rather, specific regression models for car driver injury status are needed which in addition to the risk factor also contain confounding factors as explanatory variables. The classical logistic regression model, of course, is not suitable as the injury status of the two drivers involved in the same accident can never

---

8 Maximum likelihood estimate under the fixed effects logit model (model with accident-specific parameters accounting for the pairing of units).

9 Under an unmatched design the proper estimate of the odds ratio is the so-called ‘cross-product ratio” defined as the product of the two main diagonal elements divided by the product of the two off-diagonal elements. For 2x2 tables the cross-product ratio is the maximum likelihood estimate of the odds ratio under the classical logistic regression model (model with population-averaged parameters). This model is suitable when no pairing of units is present in the data set. The latter would, for instance, be the case if for some reason only single car accidents are to be analysed. In this situation rows would correspond to car models (subject/other) and columns to driver injury status (injured/uninjured).

be regarded as two independent variables. Rather, the two-level structure of the data (level 1: accidents; level 2: accident-involved cars/drivers) must be taken into account.

Among several alternative stochastic models the fixed effects logit model appears to be most suitable for the statistical analysis of paired-by-collision car/driver data, especially when theoretical as well as practical considerations play a role. In order to obtain empirical estimates of the regression parameters and estimates of the corresponding (adjusted) odds ratios one can transform the fixed effects logit model in a certain way (called 'conditioning out accident-specific fixed effects') leading to the so-called conditional logistic regression model for matched pairs data. This model can be estimated using standard logistic regression software. Very briefly, the method for estimating the parameters associated with the categories of the risk factor car make and model and the parameters associated with the confounding factors can be described as follows:

- Eliminate all accidents from the data set where injury status of the two drivers does not differ.
- Create difference scores for all car- and driver-specific covariates (value for car 1 minus value for car 2).
- Use maximum likelihood to estimate the logistic regression predicting injury status of driver of car 1 with the difference scores as predictor variables in a model with no intercept.

Finally, it is stressed again that from paired car/driver data one cannot estimate the absolute risk of driver injury but only the comparative chance of driver injury. This, however, is completely sufficient in many situations, especially in vehicle safety studies aiming at a ranking of various car models with respect to crashworthiness.

Practical examples of driver injury odds ratio estimation based on paired-by-collision car/driver data

Description of the empirical database

Subsequently, two different car models, say car model A and car model B[10], are distinguished and driver injury status is measured by the binary variable 'driver injured yes/no'. As only car-to-car crashes are considered, each cluster consists of exactly two members (accident-involved cars). If in each two-car accident one car is arbitrarily labelled as 'car 1' and the other as 'car 2', one obtains 16 different subpopulations of accidents.

Table1[11] shows these subpopulations. The sizes of the 16 subpopulations which are also given in the table have been taken from German road traffic accident statistics 1998-2002 (car model A=Golf 2, car model B=Golf 3). As one can see, in total 3973 two-car accidents were registered by police, where a model A car collided with a model B car ('A/B crashes') or where cars of the same make and model crashed ('A/A crashes' and 'B/B crashes', respectively).

It will be shown how the driver injury odds ratio for car model A compared to car model B can be estimated taking account of the fact that the cars/drivers in the sample are paired by collision. At first, however, an analysis is presented where the clustering (pairing) of study units is erroneously ignored.

Odds ratio estimation ignoring the clustering in the sample of accident-involved cars/drivers

If the clustering in the sample of accident-involved cars is not taken into account, every single accident-involved car (there are 2x3973=7946 such cars in the sample) is treated as an individual study unit. Since for every car the binary variables 'car make and model (A/B)' and 'car driver injury (yes/no)' have been recorded, we may generate from Table 1 Table 2 as 2x2 contingency table corresponding to an unmatched cohort study design.

From Table 2 one obtains the following driver injury risk estimates for car model A and car model B (group-specific absolute risks):

$\hat{r}(A)$ = 1322/4007 = 32.99% and
$\hat{r}(B)$ = 1269/3939 = 32.22%.

Thus, the estimated relative risk of drivers of car model B (subject car model) compared to car model A (reference car model) is given by

$$\hat{\lambda} = \hat{r}(B) / \hat{r}(A) = 0.9765.$$

Instead of the two group-specific risks one could calculate the corresponding group-specific driver

[10] Car model B may not necessarily be a specific car model. Rather, it may also be interpreted as 'not car model A.

[11] In this scheme the guilty party is referred to as car 1.

No.	Make and model		Driver injury		Relevance of subpopulation for matched analysis		Number of accidents in sample
	car 1	car 2	car 1	car 2	Matched cohort design	Matched case-control design	
1	A	A	yes	yes			137
2	A	A	yes	no		X	107
3	A	A	no	yes		X	368
4	A	A	no	no			548
5	A	B	yes	yes	X		114
6	A	B	yes	no	XX	XX	77
7	A	B	no	yes	XX	XX	273
8	A	B	no	no	X		427
9	B	A	yes	yes	X		76
10	B	A	yes	no	XX	XX	46
11	B	A	no	yes	XX	XX	306
12	B	A	no	no	X		368
13	B	B	yes	yes			152
14	B	B	yes	no		X	81
15	B	B	no	yes		X	375
16	B	B	no	no			518
Total							3973
Legend: X=subpopulation of concordant pairs (not relevant), XX=subpopulation of discordant pairs (relevant)							

Table 1: Empirical database of two-car accidents

Make and model of car	Car driver injured?		Total
	yes	no	
Model A (‘other’)	1322	2685	4007
Model B (‘subject’)	1269	2670	3939
Total	2591	5355	7946

Table 2: Crosstabulation ignoring the pairing of accident-involved cars/drivers

injury odds as well to obtain the estimated driver injury odds ratio (cross-product ratio), comparing car model B to car model A:

$$\hat{\psi} = (1269/2670) / (1322/2685) = 0.9653.$$

As can be seen, both the estimated relative risk and the odds ratio estimate are very close to unity. Thus, from an analysis ignoring clustering the conclusion would be drawn that the subject car model B has the same level of crashworthiness as the reference car model A.

Statistical calculations (based on a logarithmic transformation of the two measures of comparative chance to obtain a better approximation by the normal distribution) lead to the result that the population value of the relative risk and the odds ratio does not differ from unity. For instance, the 95% confidence interval for the odds ratio ranges from 0.8789 to 1.0603.

Clearly, the above analysis would be meaningful if the m=7946 cars in the sample were independently drawn from the population of all cars that are involved in two-car crashes (A/B, A/A and B/B crashes). This, however, is obviously not the case as always pairs of cars are drawn. Consequently, any methodologically sound analysis must explicitly observe that the database actually is a sample of n= 3973 pairs of cars. If clustering is properly taken into account, it appears that the passive safety levels of the two car models under consideration are by no means identical. Rather, the crashworthiness level of car model B is significantly higher than the corresponding safety level of car model A.

Matched pairs analysis without adjustment for concomitant variables

If every matched pair of accident-involved cars is treated as a single study unit, one can build from Table 1 for each of the two possible matched study designs the corresponding 2x2 contingency table which forms the basis of a matched analysis. It becomes evident that in contrast to the above (unmatched) analysis where all subpopulations of accidents were considered, only 8 out of the 16 subpopulations of accidents are relevant in a matched analysis.

Model B car driver injured?	Model A car driver injured?		Total
	yes	no	
yes	190	319	509
no	383	795	1178
Total	573	1114	1687

Table 3: Contingency table under the matched cohort study design (Design I)

Car model of injured driver	Car model of uninjured driver		Total
	Model B	Model A	
Model B	456	319	775
Model A	383	475	858
Total	839	794	1633

Table 4: Contingency table under the case-control study design (Design II)

Let as before car model B be the subject car model. Then, under the matched cohort study design (Design I) one obtains Table 3 2x2 table.

Under the matched case-control study design (Design II) with injured drivers as cases and uninjured drivers as controls, the 2x2 table looks like Table 4.

It appears, that under Design *i* the 2x2 table contains 1687 of the 3973 accidents (i.e. 42.5% of all crashes). Under Design II the table shows the 2-dimensional distribution of 1633 accidents (or 41.1% of all crashes). Obviously, as compared with an unmatched study of two-car accidents where every single accident-involved car (there are 2x3973=7946 such cars) is treated as an individual study unit, the concept of pairing leads to smaller numbers of observations.

Although the above contingency tables for Design *i* and Design II are not identical, exactly the same conclusions about the relative chance of driver injury can be drawn from the two tables. The main result is obtained by calculating the matched odds ratio, i.e. the estimate of the odds ratio in the case of paired data. According to a well-known theorem (sometimes called the fundamental theorem of epidemiology) the matched odds ratio takes on the same numerical value under both matched designs:

$$\hat{\psi}_m = 319 / 383 = 0.8329.$$

Since 1 – 0.8329=0.1671, the main conclusion is that being the driver of a model B car will reduce the chance of injury by approximately 17% compared to a model A car. As an appropriate statistical test (McNemar's test) shows, the null hypothesis of 'no association between car make and model and car driver injury' can be rejected at the 2 percent level.

Thus, it becomes evident that when pairing of accident-involved cars is properly taken into account the superiority of car model B compared to car model A is statistically proven. If pairing is erroneously ignored, one does not come to this conclusion. Obviously, the disregard of the multi-level data structure leads to biased vehicle crashworthiness assessment results.

Matched pairs analysis with adjustment for concomitant variables

In the case of paired car/driver data, the driver injury odds ratio can be adjusted for concomitant variables like gender of car driver and/or mass of opponent car by using fixed effects regression models. If, for instance, the driver's injury status (driver injured yes/no) is assumed to be a function of the car model driven and the driver's gender, one can estimate an appropriate fixed effects logistic regression model. As a result one obtains the 'gender-adjusted' driver injury odds ratio estimate for the subject car model compared to the reference car model. The fixed effects logistic regression model may also be used to adjust the driver injury odds ratio simultaneously for several concomitant variables (e.g. driver injury odds ratio adjusted for driver age and opponent vehicle mass). Similarly, more than two car model categories can be considered.

The fixed effects logit model has been applied to empirical two-car accident data from the German traffic accident statistics. In the results reported below driver injury status is a binary variable defined as 'severely injured or killed yes/no'. To keep the example simple, the risk factor car make and model was also considered to be a binary variable defined as 'Golf-3 yes/no'. The odds ratio for the risk factor has been estimated both without and with adjustment for confounding factors using the LOGISTIC procedure of the SAS system. To illustrate the application of the fixed effects logit model a vehicle characteristic (opponent car mass in kg) and a driver characteristic (driver gender) have been selected as covariates from a larger set of possible confounders. For details see [1].

Based on a sample of n=27250 two-car accidents where exactly one driver was injured (matched case-control design) the unadjusted odds ratio for covariate 'Golf-3 yes/no' was estimated at[12]

$$\hat{\psi}_m = 1246/1576 = 0.791$$

with 95 percent confidence limits 0.734 and 0.852 (so-called Wald confidence limits). This means that compared to other cars the driver injury odds for Golf-3 is about 21 percent lower (1–0.791=0.209). In view of this result Golf-3 can be assessed as 'significantly safer than other cars'.

As Table 5 shows, the estimated odds ratio for the risk factor car make and model decreases if in addition to car make and model other covariates are included in the fixed effects logit model. The difference between the unadjusted (regression model M1) and the adjusted odds ratio (regression models M2 and M3), however, is not statistically significant since the corresponding confidence intervals overlap.

The estimation results for the fixed effects logit model M3 yield an adjusted odds ratio for the risk factor of 0.707. This can be interpreted as follows: (i) given that the opponent car has the same mass as the Golf-3 and (ii) given that the two car drivers have the same gender, the chance of being injured is for Golf-3 drivers 29 percent lower (1–0.707= 0.293) than for drivers of other car models. In M3 the upper confidence interval limit (UCL) is far below unity (0.773). Therefore, the crash performance of the Golf-3 can be considered as significantly better than the performance of other car models. For each of the three regression models the confidence interval for the odds ratio related to the risk factor is given in Table 5. As the three confidence intervals overlap, the crashworthiness assessment of Golf-3 does not substantially change after adjustment for the factors 'driver gender' and 'mass of opponent car'.

There may, however, be covariates where the unadjusted and adjusted odds ratio differ significantly. Among other things Table 5 shows that the estimated odds ratio for the covariate 'driver gender' equals 0.480[13] if one only adjusts for car make and model (regression model M2), but is equal to 0.539 if, in addition, adjustment is made

Fixed effects logit model of diver injury					
Model	Covariate(s)	Odds ratio for covariate			Relative length of CI[1)]
		LCL	Estimate	UCL	
M1	Golf-3 yes/no	.734	.791	.852	.149
M2	Golf-3 yes/no	.700	.756	.816	.153
	Male driver yes/no	.462	.480	.498	.075
M3	Golf-3 yes/no	.647	.707	.773	.178
	Male driver yes/no	.515	.539	.565	.093
	Mass of opponent car (100 kg)	1.277	1.289	1.302	.019
1) Relative length of confidence interval=(UCL–LCL)/Estimate					

Table 5: Odds ratio estimates obtained from the fixed effects logit model

also for mass of opponent car (regression model M3). Looking at the confidence intervals it can be concluded that the adjusted odds ratio is different from the unadjusted. The absolute difference between the two values is, however, rather small (0.539–0.480=0.059). As driver gender and opponent car mass are largely independent determinants of driver injury status, this result is not surprising.

Finally, it should be noted that in the regression models M2 and M3 the coefficients and thus the odds ratios for the various determinants of driver injury status are estimated with quite different levels of accuracy. As the relative lengths of the confidence intervals (length of interval divided by midpoint of interval) show, the estimation of the coefficient for mass of opponent car is by far the most accurate. According to M3, it is almost certain (confidence level 95 percent) that colliding with a 'heavier' car rather than with a 'lighter' car (mass difference 100 kg) increases the driver's odds of being injured between 27.7 and 30.2 percent.

Alternative statistical models of driver injury for paired-by-collision car/driver data

In empirical crashworthiness studies where data on cars/drivers matched in pairs are to be analysed and where car driver injury status is the criterion variable of interest, the fixed effects logit model is certainly an appropriate statistical tool. This is because the model

- has a sound theoretical basis,
- is relatively easy to understand and to handle also for non-statisticians and
- can be estimated using standard statistical software.

[12] In this example Golf-3 is compared with all other car models, not only with Golf-2 as was the case in the previous example.

[13] Being a male driver reduces injury odds by 52% compared to female drivers.

As can be expected, however, various alternatives to the fixed effects logit model are available. The choice between these alternative models mainly depends on the scaling of the dependent variable car driver injury status. Subsequently, some of the candidate regression models are mentioned briefly without going into any methodological detail.

Driver injury status as a binary variable

When driver injury status is a binary variable (injury yes/no) random effects models of the logit and probit type can be considered as alternatives to the fixed effects logit model. See [4], p. 837-849, and [8], p. 62-70. Another alternative to be mentioned is the bivariate logit model with covariates describing not only the accident but also the two cars and drivers involved in the accident. This model was first proposed by [9]. It should be noted here that bivariate modelling approaches where only accident-specific (but no car- and driver-specific) covariates can be incorporated, are not really useful for crashworthiness assessment purposes. This is the reason why, for instance, the bivariate logit model derived from the log-linear model (see, for instance, [10], p. 223-225) is not suitable in our context.

Finally, the Bradley-Terry model developed for paired comparisons is another candidate modelling approach. See [4], p. 270-276, [11], p. 102-103, and [12]. In the SARAC 2 project this model has been applied to two-car accident data [13] .

Driver injury status as a variable with several ordered levels

When driver injury status is measured on an ordinal scale (e.g. uninjured, slightly injured, severely injured, killed or, alternatively, AIS 0 to 6) the fixed effects cumulative logit model can be applied to assess the role of the various car and driver characteristics as determinants of driver injury. For a software-oriented description see [8], p. 70-74.

As can be expected, the proper application of the various models mentioned above requires deeper knowledge of statistical theory and more specialised software.

References

1 H. HAUTZINGER: Design and Analysis of Matched Studies in Empirical Car Safety Research, Report S2-246, SARAC 2 project, Heilbronn/Germany, IVT, 2006

2 M. WOODWARD: Epidemiology – Study Design and Data Analysis, Boca Raton/ London/New York/Washington DC, Chapman & Hall/CRC, 1999

3 H. GOLDSTEIN: Multilevel Statistical Models, 2nd edn., London, Arnold, 1995

4 W.G. GREENE: Econometric Analysis, 4th edn., Upper Saddle River, NJ, Prentice Hall, 2000

5 H. HAUTZINGER, D. HEIDEMANN: Econometric Analysis of the German Mobility Panel, in Travel Behaviour Research – The Leading Edge (ed. D HENSHER), London, Pergamon, 2000, p. 481-499

6 H. HAUTZINGER: Description and Analysis of Existing Car Safety Rating Methods, Report of Sub-Task 1.1 of SARAC 1 project, Heilbronn/Germany, IVT, 2001

7 J.-L. MARTIN, Y. DERRIEN, B. LAUMON: Estimating relative driver fatality and injury risk according to some characteristics of cars and drivers using matched-pair multivariate analysis, Paper presented at the ESV Conference 2003

8 P.D. ALLISON: Fixed Effects Regression Methods for Longitudinal Data Using SAS, Cary, NC, SAS Institute Inc., 2005

9 D.R. COX: The Analysis of Multivariate Binary Data, Applied Statistics (Journal of the Royal Statitistical Society, Series C), 21 (2), 1972, p. 113-120

10 P. McCULLAGH, J.A. NELDER: Generalized Linear Models, 2nd edn., London, Chapman & Hall, 1992

11 D.R. COX: Analysis of Binary Data, London, Chapman & Hall, 1970

12 D.R. HUNTER: MM Algorithms for Generalized Bradley-Terry Models, The Annals of Statistics, 32 (1), 2004, p. 384-406

13 C. PASTOR: Correlation of NCAP Star Rating and German Accident Data Using Methods of ‘Paired Comparisons’, internal paper, SARAC 2 project, Bergisch Gladbach, BASt, 2004