PERGAMON

Accident Analysis and Prevention 35 (2003) 537–547

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Comparison of road crashes incidence and severity between some French counties

E. Amoros*, J.L. Martin, B. Laumon

Transport, Work and Environment Epidemiology Laboratory, French Research Institute on Transport and Safety (INRETS), 25 Avenue F. Mitterrand, Case 24, 69675 Bron Cedex, France

Received 15 February 2001; received in revised form 12 February 2002; accepted 22 February 2002

Abstract

Our aim is to compare traffic safety among several counties in France, and explore whether observed differences can be explained by differences in road types distribution and by differences in socio-economic characteristics between counties. Traffic safety is measured by incidence and severity, where incidence is defined by the ratio of counts of injury accidents and exposure, measured by the amount of kilometres driven. Severity is measured by the ratio between fatal and injury accidents. These indexes are analysed in the framework of Generalised Linear Models: counts of injury accidents are analysed with a Negative Binomial regression, which accounts for over-dispersion. Severity being the proportion of fatal accidents among injury accidents corresponds to the probability of a Binomial setting and this is modelled by a logistic regression.

This modelling provides an easy way to adjust for covariates such as road type, environment (urban/rural) and evolution over time, and to test their possible interactions. We find that the time trend of each indice (incidence and severity) is the same across counties and across road types. There is a significant interaction between county and road type, meaning that, first, differences in traffic safety between counties are not fully explained by different road type distributions, and second, that the "ranking" of counties in term of incidence or severity varies according to the road type considered, and vice-versa. It was planned to explore global characteristics of the counties (driving and socio-economic data) as possible explanatory factors of differences between counties, but the existence of an interaction of county with road types shows the necessity of collecting and exploring characteristics of the sub-levels of road type within county.

Keywords: Aggregated data; County; Negative binomial; Logistic; Over-dispersion; Interaction

1. Introduction

Safety indexes differ between countries, and it is useful to analyse these differences in order to obtain a better understanding of the factors related to crashes (Brühning and Alevisos, 1991). However, such an analysis is difficult because of differences between countries in the definitions of crash and socio-economic data, e.g. severe/slightly injured persons, alcohol consumption; furthermore data collection methods may differ. This is not the case when comparing areas within one country. Taking advantage of this, comparison is made between some areas in France. There are geographical and administrative units, the counties (called "départements" in French), which show differences in road safety. These counties are in charge of managing a network of roads which are known as the "county roads". Some minor safety decisions about these county roads can be taken at the county level.

The aim of this study is to analyse differences in traffic safety between counties, in order to better identify possible risk factors of traffic safety. We explore whether observed differences can be explained by two kinds of factors: some related to distribution of road types within counties, and some that characterise the counties: driving and socio-economic factors.

We use aggregated data, and we focus on eight counties, that constitute the Rhône-Alpes region in south-east France. Traffic safety is measured by (i) incidence and (ii) severity of crashes involving casualties (killed or injured persons). Incidence is defined by the ratio between counts of injury accidents and exposure, measured by the amount of kilometres driven. Severity is measured by the ratio between fatal accidents and injury accidents. The analysis focuses on the comparison of the counties. Can differences in traffic safety between counties be explained by differences in road

* Corresponding author. Present address: Institute of Cancer Epidemiology, Danish Cancer Society, Straudbovlevarden49, 2100 Copenhagen, Denmark. Tel.: +45-35-25-76-13; fax: +45-35-25-77-31.

E-mail address: amoros@cancer.dk (E. Amoros).

0001-4575/02/$ – see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0001-4575(02)00031-3

type distribution, and in urban/rural road distribution? It is indeed known that frequency of crashes, as well as their severity depend on the road type and on the urban/rural area. This is addressed in the modelling by including the categorical variables county and road type. The modelling also provides an easy way to test for possible interaction and take it into account, if significant. Calendar time is also included in the model, as well as its possible interaction with other covariates.

We also explore socio-economic variables as possible risk factors for traffic safety. Exploration of the link between socio-economic characteristics of counties (i.e. of their population) and road crashes makes sense as people tend to have road crashes in the county where they live (Laumon et al., 1997).

2. Data

The Rhône-Alpes region is in the south-east of France, and it is divided into eight counties: Ain, Ardèche, Drôme, Isère, Loire, Rhône, Savoie, Haute-Savoie (Fig. 1). This region has a total population of 5.5 million (from 282,000 in Ardèche to 1,540,000 in the Rhône county), and a total area of 43,700 km^2 (from 3250 km^2 in the Rhône to 7430 km^2 in Isère).

2.1. Road type

There are four main road types in France. These are motorways, national roads, county roads, and local roads. They differ by the level at which they are administratively managed. Motorways are managed by both the state and companies, national roads by the state, county roads by the county, local roads by cities/villages. Usually, they also differ by the amount of traffic: traffic decreases with level of network.

Fig. 1. Map of France, Rhône-Alpes region, and its eight counties.

Roads are also classified into urban/rural area: a road (or part of road) is defined as urban when it lies within the area delimited by the roadsigns indicating beginning and end of city/village.

2.2. Crash data

Available road crash data include counts of crashes involving casualties, i.e. injury accidents, counts of crashes involving fatalities, i.e. fatal accidents, as well as counts of fatalities, of severely injured people, and of slightly injured people. Fatalities are defined as deaths occurring within 6 days after the accident; severely (respectively slightly) injured people are those requiring more than (respectively less or equal) 6 days of hospital care. These data are provided by the Regional Observatory of Road Safety, at the Ministry of Transport; they come from police reports. Counts of property-damage only accidents are not available.

Data are available by county, year, road type and urban/rural roads. This means that the level of aggregation of accident data is road type within county.

2.3. Exposure

At the county and year level, possible measures of exposure to road crashes are: number of registered vehicles, petrol (gas) consumption, total population, and urban population. However, exposure has to be measured at the road type level (in order to adjust for it). At this level, available data are counts of vehicles on motorways and on national roads, as well as the length of the motorway network, of the national roads network, and of the combined county and local roads network.

2.4. Characteristics of the counties

The counties can be characterised by some global variables, i.e. summary variables defined at the level of the counties. Variables considered here are those available by county (and possibly by year), for which there is some assumption about possible link with traffic safety.

Socio-economic data considered here are: distribution of population by age class (Mercer, 1987), by matrimonial status (Mannering, 1993), number of unemployed people (total, and <25 years old) (Mercer, 1987), amount of income tax (Baker et al., 1987), number of suicides, and of drug offences. Data about driving are: number of driving licenses delivered (Fridstrøm and Ingebrigtsen, 1991), number of accompanied learner drivers, and number of driving offences. Some other data, which may play a role in the frequency and/or severity of crashes were not considered as their distribution (and possible trend) is thought to be the same in all counties: this is the case of the distribution of population by sex for instance; however, distribution of drivers by sex could be interesting (Mercer, 1987; Mannering, 1993), but is not available.

We mentioned that some safety decisions could be made about county roads. Unfortunately, information on those is not available. On the other hand, these are only minor decisions (for instance improving roadsigns on some county roads); major decisions (such as speed limits, fines, seat belt, etc.) are taken on the state level, so that it would not affect the comparison between counties.

The annual data were only available for the period 1986–1993 (up to 1994 for the crash data) at the time the study was done.

3. Safety indexes

To study traffic safety, some choice must be made about what to measure and how. What is of public health concern is the number of casualties (killed and injured people), and this is the result of the frequency of crashes and of their severity, and more precisely, the product of these two factors. A decrease in the number of casualties can, therefore, be obtained by a decrease of one of these two factors or by both; we refer to the decomposition method (Hemenway, 1998; Li et al., 1998). To put it in another way, the probability of casualties is the probability that occupants are killed or injured given that a crash occurred (*conditionally* on the accident), multiplied by the probability of the crash itself.

Unfortunately, counts of property-damage only crashes are not available. Only counts of crashes involving casualties are available. This means that we can not study the frequency of any crash, but only the frequency of a crash involving casualties. It also means that we can only analyse severity as the proportion of the most severe crashes among injury crashes, and not the proportion of injury accidents among all accidents (including property-damage only accidents).

The unit considered here is the accident and not the person, even though it is the number of casualties that is of public health concern. This is because we expect similar results and interpretations on the risk factors studied here whatever unit we use, and because the casualties are more complex to model as they are not independent of one another (as one accident can lead to several casualties) whereas accidents can be considered independent of one another.

3.1. Definition of incidence

In order to define the measure of crash frequency, the epidemiological notion of incidence is used here. The annual incidence rate for a particular calendar year is the number of new cases occurring during that year, divided by the number of units exposed and prone to become a case, and multiplied by their duration of exposure during that year (Breslow and Day, 1987). In an epidemiological framework, a case is the occurrence of death or disease; here, a case is the occurrence of an accident involving casualties. As for the denominator in the definition of the incidence, i.e. the exposure to the risk of accident, it can be merely measured by the number of inhabitants or by the number of vehicles, but this has many drawbacks; the main one is that the different amounts of exposure per inhabitant or per vehicle are not taken into consideration (Andreassen, 1991). Therefore, the exposure to road crashes is usually estimated by the number of kilometres driven (Mercer, 1987), or, rather rarely, by the time spent driving (Chipman et al., 1992; Joly et al., 1991). In our study, the choice of a measure of exposure is moreover driven by the fact that we need to measure exposure for each road type, so that we can adjust for road type in the analysis. On motorways and on national roads, counts of vehicles by small section are available, together with length of the network so that it is possible to estimate the number of kilometres driven on motorways and on national roads (separately). Counts of vehicles are not available on the other road types, but an estimation of the number of kilometres driven on county roads and local roads combined (in each county) can be obtained by the difference between the total number of kilometres driven in the county minus the number of kilometres driven on motorways and on national roads. The total number of kilometres driven in the county is estimated by the amount of petrol sold in the county applied to the average consumption by vehicle type, and taking account of the distribution of the vehicles types of the county. This is an approximation, as petrol sold in one county can be used to travel to/in another one (but this is the only possibility left). The study is, therefore, limited to the incidence of crashes using a grouping of road types in three categories (motorways/national roads/county and local roads) instead of the four existing ones. The study period is 1986–1993.

3.2. Definition of severity

As discussed above, the severity index will measure the degree of severity of injury crashes, and it will use the accident as the unit, not the person. Therefore, we define the severity index by the ratio between counts of fatal accidents and counts of injury accidents.

This index is defined at a more acute level than the incidence index, in the sense that corresponding crash data are available on each of the four existing categories of road types, and simultaneously by urban/rural categorisation. However, the analysis is based on six categories instead of the eight combinations. There are: motorways/national roads/county roads, rural/county roads, urban/local roads, rural/local roads, urban. We have grouped together urban and rural motorways as it does not make sense to dissociate them: same infrastructure, often same speed limit. Urban national roads and rural national roads have also been grouped together, as infrastructure and speed limit are quite often the same, and because we are concerned about misclassification between the two. The study period is 1986–1994 for this severity index.

4. Method

4.1. Statistical distribution of the safety indices

The incidence has been defined as the ratio of counts of crashes involving casualties and the amount of kilometres driven. What we are interested in modelling is the occurrence of injury crashes conditional on the exposure, i.e. the exposure is considered as known.

Counts of car crashes are assumed to have a Poisson distribution (Greenwood and Yule, 1920; Kerrich, 1951), such that the probability of m crashes in year t, in county i, on road type k, is: $P[Y(t, i, k) = m] = e^{-\lambda_{ikt}}(\lambda_{ikt})^m/m!$ where the mean parameter λ_{ikt} is the expected number of crashes.

Such data can be analysed with a Poisson regression, or with a Negative Binomial regression (see below).

The severity index has been defined as the ratio of counts of fatal accidents divided by counts of injury accidents. This corresponds to a Binomial setting where the trial is the occurrence of an injury accident, and the event is a fatal accident. The probability of a fatal accident can be modelled with a logistic regression.

These three types of regressions belong to the Generalised Linear Models.

4.2. Generalised Linear Models

Generalised Linear Models (McCullagh and Nelder, 1989) are defined by three characteristics which are (*i*) the random component (Y) has a distribution within the exponential family, (ii) its mean is linked to a linear combination of systematic components (covariates) through (iii) a monotonic differentiable function (g, so-called link function):

$$g(E(Y)) = g(\lambda) = \eta = \sum \beta_j x_j$$

In the case of Poisson regression, the model is: $\log(E(Y)) = \log(\lambda) = \eta = \sum \beta_j x_j$, where λ is the mean parameter of the Poisson distribution.

In the case of logistic regression, the model is: $\text{logit}\,(E(Y)) = \text{logit}\,(p) = \eta = \sum \beta_j x_j$, where p is the probability parameter of the Binomial distribution.

4.3. Over-dispersion

The Poisson distribution has the property that the variance is equal to the expected value, i.e. $\text{var}(Y) = \lambda = E(Y)$. This does not usually hold in practice. Most of the time, observations have a greater variance than their mean. This over-dispersion can be taken into account in several ways, the most common being quasi-Poisson and negative binomial models (Aitkin et al., 1989).

In the quasi-Poisson approach, the modelling of Poisson is changed by allowing the variance to be greater than the mean parameter in a linear way, i.e. $\text{var}(Y) = \phi\lambda$, with $\phi > 1$. The fitting of the model is done in a similar way, resolving so-called quasi-score equations. The parameter estimates are the same as when there is no over-dispersion; only the confidence intervals are wider as the variance is inflated by ϕ. The estimation of ϕ can be done by dividing the deviance by the degrees of freedom or, as suggested by McCullagh and Nelder (1989), by dividing Pearson Chi-square by its degrees of freedom.

Another approach about over-dispersion is to assume that the mean parameter λ is itself random, and it is typically assumed that it follows a Gamma distribution. This distribution ensures positive values for λ, and is rather flexible (King, 1989). Conditional on λ, the distribution of accidents is a Poisson one, and the marginal distribution has been shown to be a Negative Binomial (Gourieroux and Montfort, 1984a,b).

If the gamma distribution has shape parameter b (i.e $E(\lambda) = \text{var}(\lambda) = b$), then the negative binomial distribution will be characterised by $\text{var}(Y) = \mu + k\mu^2$, with $\mu = E(Y)$ and $k = 1/b$, k corresponding to the over-dispersion parameter.

In the binomial setting, some over-dispersion can also be allowed for. In the same way as the quasi-Poisson, the variance function $V(\mu) = \mu(1 - \mu)$ is extended to $V(\mu) = \phi\mu(1 - \mu)$ where ϕ is the over-dispersion parameter. The estimation of the model is also done in the same way as in the quasi-Poisson; in particular, the coefficient of over-dispersion is estimated by dividing Pearson Chi-square by its degrees of freedom.

4.4. Details about the modelling (offset)

As we are interested in comparing the incidence and severity between the different counties, both being defined as ratios, we treat the denominator as an offset in the modelling, i.e. its coefficient is fixed to 1. In the modelling of the incidence index, this means that the exposure, i.e. the amount of kilometres driven is explicitly specified as an offset. In the modelling of severity, the ratio corresponds to the probability parameter of a Binomial, and it is modelled as such in the logistic regression (which is equivalent to treating the denominator as an offset).

4.5. Model validation

Several tests are used here to check the validity of the models fitted. First, the assumed distribution of the response variable is checked. This is done by testing the normality of the deviance residuals, using the Shapiro–Wilk test (Shapiro and Wilk, 1965). Secondly, the variance function is checked, using a graphical test that plots the absolute residuals (on the transformed scale) versus the fitted values (McCullagh and Nelder, 1989). If the modelled variance function does not increase rapidly enough with $\hat{\lambda}$, the scatter of residuals will show an increasing trend. For instance, if the variance function was linear in λ, a quadratic function in λ should be tried. A decreasing trend would indicate the contrary.

Thirdly, the canonical link function g is checked, that is, whether the transformed expected value of the dependent variable—through the link g—is a linear function of the covariates. A graphical test (McCullagh and Nelder, 1989) is used, that plots the adjusted dependent variable versus the fitted values; a straight line should be obtained.

The fitting of the models is done using the softwares SAS (GENMOD procedure) and Stata (nbreg function).

4.6. Model selection, accounting for road type, and environment

The idea is to explore whether differences in traffic safety between counties remain after adjusting for road type, environment and time trend. We include the county variable in the model and we are interested in seeing whether relative risks (RR) associated to the different counties remain significantly different when adjusting on the infrastructure and time variables. As mentioned, the road type (and environment when available) variable is included in the model in order to adjust on it (without testing it), as it is known that incidence and severity are different among road types (and that the distribution of road types is different between counties); incidence and severity are also different between urban and rural roads (and the distribution of urban/rural roads is different between counties). It is reminded that a grouping of road types into three categories is used when analysing incidence (motorways/national roads/county and local road); when analysing severity, the four existing road types are combined with environment into six categories (motorways/national roads/rural county roads/urban county roads/rural local roads/urban local roads). The calendar time variable is then tested for inclusion, and whether it should be included as a categorical variable or as a continuous one. If continuous, we assume a linear trend on the transformed scale (corresponding to an exponential trend in the indexes themselves). A continuous coding of calendar time is preferred over a categorical one, as the plots of traffic safety indexes do not show complex trends over time.

Also, possible interactions between the covariates are then investigated, as well as order 2 interaction between time, road type and county, if previous order 1 interactions are significant. It is emphasised that the easy exploration of interactions is one asset of this study, and that their meaning are of real interest in the interpretation of the results.

4.7. Model selection, exploring global characteristics of the counties

We want to explore whether some global characteristics (driving and socio-economic data) of the counties could explain differences in traffic safety between counties (after adjusting on road type). In term of modelling, instead of including the categorical county variable, we test these variables characterising the counties for inclusion in the model, i.e. we test which of these can explain the (systematic part of) variability of the crash data. The road types and calendar time variables are included in the model (without testing them) in order to adjust for them.

In all model selections, we allow for over-dispersion in the model from the start of the ascending selection process. Over-dispersion is a common phenomena in crash data, and we are concerned that standard errors might be inappropriately small if we do not allow for over-dispersion in the model.

4.8. Parameter interpretation

The analysis of the incidence index with a Poisson or Negative Binomial regression allows the estimation of Relative Risks (RR), which are here equal to the ratio λ_j/λ_0 where λ_j is the incidence of category j, and λ_0 corresponds to a chosen reference. If the model is merely: $\log(\lambda_{ik}) = c + \beta_i\,\text{county}\,i + \alpha_k\,\text{road type}\,k$, then $\exp(\beta_i)$ gives the relative risk of county i compared to the county of reference, adjusted for road type.

If the model includes interaction such as: $\log(\lambda_{ik}) = c + \beta_i\,\text{county}\,i + \alpha_k\,\text{road type}\,k + \gamma_{ik}\,\text{county}\,i \times \text{road type}\,k$, then the relative risk of county i compared to the county of reference depends on the road type: on road type k, the relative risk between county i and the reference county is $RR_{i(k)} = \exp(\beta_i + \gamma_{ik})$.

The analysis of the severity index allows the estimation of odds ratios (OR), i.e. the ratio of odds of a category j and a category of reference. This is because logit(p) is modelled. For instance, if the model includes county and road type, such as: $\text{logit}\,(p_{ik}) = \log\,(p_{ik}/1 - p_{ik}) = c + \beta_i\,\text{county}\,i + \alpha_k\,\text{road type}\,k$, then we obtain: $\exp(\beta_i) = \text{OR}_i = (p_i/1 - p_i)/(p_0/1 - p_0)$. In practice, if the risk is small, the odds ratio is a good approximation of the relative risk.

5. Results

Relative Risks are expressed relatively to a reference within a group, and this reference is usually chosen for its lowest risk. The chosen reference county is the Rhône, as this is the most urbanised one, and severity is usually lower in urban areas than in rural ones. Concerning road types, national roads are chosen as the reference as the other types are not appropriate (no motorway in Ardèche, county and local roads considered either separately or combined). Also, the Rhône county has the highest number of crashes by far, and this should ensures stability of the estimates. These county and road types references may not be the ones with the lowest risk when studying incidence, but it is clearer to keep the same ones.

5.1. Incidence of injury crashes

The number of events on which the study of incidence is based is given in Table 1: counts of road crashes involving

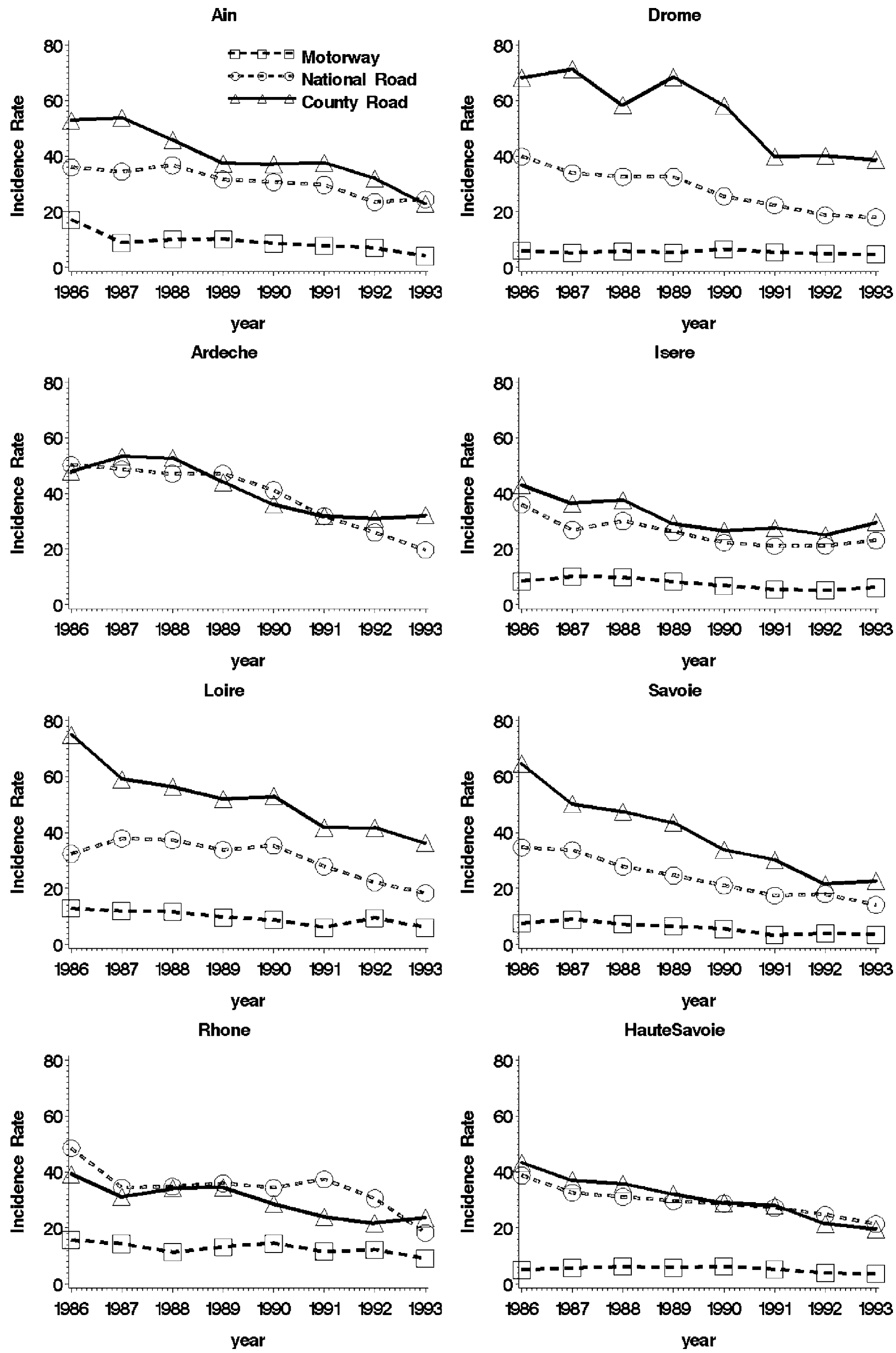

Fig. 2. Incidence of accidents involving casualties, by road type and county.

casualties are displayed by road type, over the period 1986–1993. The average incidence rate of road crashes in the Rhône-Alpes region in the period 1986–1993 is 28.5 per 100 millions kilometres driven. It ranges from 17.2 on motorways, to 50.2 on national roads, and up to 74.9 on other roads (county and local roads). Fig. 2 shows the incidence for the three road types, on each county, over the period 1986–1993. There seems to be a decreasing trend on all road types, in all counties. Globally, the levels of incidence seem to differ between counties. Also, the pattern of incidence levels of the three road types is different between counties: in some counties the incidence rates on motorways and national roads are quite similar whereas they are quite different in other counties.

Table 2 shows the distribution of road types by county, in term of kilometres, and in term of amount of vehicles flows. For instance, proportion of kilometres travelled on motorways can be as low as 14% of the traffic in a given county and as high as 43%.

Table 1
Total numbers of accidents involving casualties, over the period 1986–1993

County	Road crashes involving casualties		
	Motorways	National roads	County and local roads
Rhône	2503	26906	3973
Ain	405	7012	2608
Ardèche	–	3468	1911
Drôme	935	8406	2210
Isère	1292	10457	4351
Loire	625	11636	2697
Savoie	208	4827	2855
Haute-Savoie	288	7102	3567

As mentioned previously, we start the modelling by including the categorical variables county and road type. After forward selection, the final model includes the following variables: county, road type, year with a continuous coding, and interaction between county and road type. It is written as:

$$\log(\lambda_{ik}) = c + \beta_i \text{ county } i + \alpha_k \text{ road type } k + \rho \text{ year} + \gamma_{ik} \text{ county } i \times \text{road type } k$$

The test of a continuous coding of year versus a categorical one was significant. However, we chose the continuous coding as the categorical coding does not indicate a more complex trend than a linear one. Interaction of year with county or year with road type were not significant. The modelling with a Poisson regression is rejected by the checks of fit: the normality of the deviance residuals was rejected. We, therefore, chose to model the data with a Binomial Negative regression, as it allows for another form of over-dispersion. The same variables were included in the model during the forward selection process. This model is not rejected by the checks of fit: the test of normality of the deviance residuals has a P-value of 0.86; the graphical check of the log link function is very satisfactory, and the graphical check of the variance function is satisfactory; the dispersion parameter of the Negative Binomial is estimated at 0.01.

This absence of interaction between year and county, as well as between year and road type, means that the time trend is not different across counties, nor across road types. Incidence of injury crashes decreases in all eight counties and on all road types (RR = 0.91 between any two consecutive years). The interaction between county and road type means that we can not make a global comparison of the incidence between counties, but that the comparison can only be made within each road type. In another way, the "ranking" of counties according to their incidence varies with the road type considered. Vice-versa, the ranking of road types according to their incidence depends on the county considered; this corresponds to what can already be seen in Fig. 2.

Results of the model are shown in Table 3: RR are provided, and one can directly read the RR between counties by road type, in the interaction term. Within each road type, one reads the RR of every county compared to the reference county (Rhône). For instance, one can see that incidence of injury crashes on Drôme's county and local roads is 2.3 times higher than incidence on Rhône's similar roads, whereas incidence on Drôme's motorways is 0.54 times less than incidence on Rhône motorways.

5.2. Severity of injury crashes

The number of fatal crashes on which the analysis of severity is based on are displayed in Table 4. Counts of fatal road crashes are given, over the period 1986–1994 by road type. Average severity in the Rhône-Alpes region in the period 1986–1994 is of 7.4 fatal crashes per 100 injury crashes. Severity depends on the road type (Lassarre and Hoyau, 1995), and in the Rhône-Alpes region, the average severity on motorways is 9.0 fatal crashes per 100, on national roads 9.5, on rural county roads 11.4, on urban county roads 5.1, on rural local roads 7.7, and on urban local roads 2.4.

Distribution of traffic between (three groups of) road types was given in Table 2. Distribution of traffic between urban and rural roads is not available, but a picture of the variability among the eight counties of the Rhône-Alpes region is given by the urbanisation rates (i.e. the percentage of people living in urban areas) shown in Fig. 3; they range from 50 to 92%.

We start the modelling with the inclusion of the categorical variables county and road type (six categories). After forward selection, the final model includes the following

Table 2
Length of road network, distribution of km travelled, by county (in 1993)

County	Motorways		National roads		County and local roads		Total number of km travelled (million)
	Network (km)	Travelled (km, %)	Network (km)	Travelled (km, %)	Network (km)	Travelled (km, %)	
Rhône	160	20	219	8	12690	72	16282
Ain	170	21	359	19	17530	60	5304
Ardèche	0	–	296	35	16270	65	1941
Drôme	134	43	205	20	13530	37	5564
Isère	242	30	591	24	20000	46	8792
Loire	120	20	293	22	16460	58	5400
Savoie	95	15	421	38	7030	47	4240
Haute-Savoie	145	14	396	26	10870	60	6130

Table 3
Incidence of road crashes and estimates of the model

Variable	Estimate	S.E.	*P*	RR
Intercept	−14.58	0.04	<0.001	
Road type (for the county of reference: Rhône)				
Motorway	−0.97	0.06	<0.001	0.38
County and local road	−0.14	0.05	0.007	0.87
National road	[–]	[–]	[–]	1.00
Year	−0.09	0.004	<0.001	0.91
Interaction (between road type and county)				
National road × Ain	−0.10	0.06	0.081	0.90
National road × Ardèche	0.11	0.06	0.053	1.12
National road × Drôme	−0.22	0.06	<0.001	0.80
National road × Isère	−0.27	0.05	<0.001	0.76
National road × Loire	−0.12	0.06	0.039	0.89
National road × Savoie	−0.38	0.06	<0.001	0.68
National road × Haute-Savoie	−0.15	0.05	0.006	0.86
National road × Rhône	[–]	[–]	[–]	1.00
Motorway × Ain	−0.30	0.09	<0.001	0.74
Motorway × Ardèche	–	–	–	–
Motorway × Drôme	−0.61	0.08	<0.001	0.54
Motorway × Isère	−0.27	0.08	<0.001	0.76
Motorway × Loire	−0.19	0.09	0.029	0.83
Motorway × Savoie	−0.45	0.10	<0.001	0.64
Motorway × Haute-Savoie	−0.73	0.10	<0.001	0.48
Motorway × Rhône	[–]	[–]	[–]	1.00
County and local road × Ain	0.39	0.08	<0.001	1.48
County and local road × Ardèche	0.21	0.08	0.007	1.23
County and local road × Drôme	0.83	0.08	<0.001	2.29
County and local road × Isère	0.35	0.07	<0.001	1.42
County and local road × Loire	0.67	0.08	<0.001	1.95
County and local road × Savoie	0.62	0.08	<0.001	1.86
County and local road × Haute-Savoie	0.18	0.08	0.018	1.20
County and local road × Rhône	[–]	[–]	[–]	1.00
Over-dispersion	0.01	–	–	

covariates: county, road type, year as a continuous variable, and interaction between county and road type. It is written as:

$$\text{logit}\,(p_{ik}) = c + \beta_i\,\text{county}\,i + \alpha_k\,\text{road type}\,k + \rho\,\text{year} + \gamma_{ik}\,\text{county}\,i \times \text{road type}\,k$$

The variable year was kept with a continuous coding as the test of a categorical versus a continuous coding was not significant, and a continuous coding leads to a more parsimonious model. No interaction of year with county, or year with road type was found. The model is not rejected by the checks of fit: the test of Normality of the deviance residuals a *P*-value of 0.49; the graph of the variance function is satisfactory, as well as the graph of the link function.

The absence of interaction between year and county means that the severity trend is similar across counties; the absence of interaction between year and road type

Table 4
Total numbers of fatal accidents over the period 1986–1994

County	Motorways	National roads	County roads, rural	County roads, urban	Local roads, rural	Local roads, urban
Rhône	140	295	294	282	20	370
Ain	33	337	437	139	28	53
Ardèche	0	158	169	39	21	19
Drôme	137	243	319	94	30	58
Isère	125	479	459	199	39	109
Loire	41	251	230	141	28	119
Savoie	19	327	173	58	33	36
Haute-Savoie	44	393	280	123	40	109

Table 5
Severity of road crashes and estimates of the model

Variable	Estimate	S.E.	*P*	Odds ratio
Intercept	−2.69	0.07	<0.0001	0.07
Road type (for the county of reference: Rhône)				
Motorway × Rhône	−0.31	0.11	0.004	0.73
Local road (urban)	−1.37	0.08	<0.0001	0.25
Local road (rural)	−0.42	0.25	0.09	0.66
County road (urban)	−0.46	0.09	<0.0001	0.63
County road (rural)	0.67	0.09	<0.0001	1.95
National road	[–]	[–]	[–]	1.00
Year	0.02	0.005	0.0001	1.02
Interaction (between road type and county)				
National road × Ain	0.61	0.09	<0.0001	1.84
National road × Ardèche	0.13	0.11	0.24	1.14
National road × Drôme	0.44	0.09	<0.0001	1.55
National road × Isère	0.40	0.08	<0.0001	1.49
National road × Loire	0.27	0.09	0.003	1.31
National road × Savoie	0.49	0.09	<0.0001	1.64
National road × Haute-Savoie	0.42	0.08	<0.0001	1.53
National road × Rhône	[–]	[–]	[–]	1.00
Motorway × Ain	−0.24	0.23	0.30	0.79
Motorway × Ardèche	–	–	–	1.00
Motorway × Drôme	0.63	0.16	0.0001	1.87
Motorway × Isère	0.18	0.16	0.26	1.19
Motorway × Loire	−0.08	0.21	0.70	0.92
Motorway × Savoie	0.02	0.28	0.94	1.02
Motorway × Haute-Savoie	0.66	0.21	0.002	1.94
Motorway × Rhône	[–]	[–]	[–]	1.00
Local road (urban) × Ain	0.004	0.18	0.98	1.00
Local road (urban) × Ardèche	0.10	0.27	0.70	1.11
Local road (urban) × Drôme	−0.46	0.18	0.01	0.63
Local road (urban) × Isère	−0.02	0.14	0.88	0.98
Local road (urban) × Loire	−0.33	0.15	0.02	0.72
Local road (urban) × Savoie	−0.40	0.20	0.05	0.67
Local road (urban) × Haute-Savoie	0.16	0.14	0.25	1.18
Local road (urban) × Rhône	[–]	[–]	[–]	1.00
Local road (rural) × Ain	−0.21	0.33	0.53	0.81
Local road (rural) × Ardèche	0.45	0.35	0.21	1.57
Local road (rural) × Drôme	−0.17	0.32	0.60	0.84
Local road (rural) × Isère	0.18	0.31	0.56	1.20
Local road (rural) × Loire	0.40	0.33	0.22	1.50
Local road (rural) × Savoie	0.39	0.32	0.22	1.48
Local road (rural) × Haute-Savoie	0.46	0.31	0.14	1.58
Local road (rural) × Rhône	[–]	[–]	[–]	1.00
County road (urban) × Ain	−0.20	0.14	0.16	0.82
County road (urban) × Ardèche	−0.14	0.21	0.50	0.87
County road (urban) × Drôme	−0.63	0.16	<0.0001	0.53
County road (urban) × Isère	−0.14	0.13	0.27	0.87
County road (urban) × Loire	−0.41	0.14	0.005	0.67
County road (urban) × Savoie	−0.53	0.18	0.003	0.59
County road (urban) × Haute-Savoie	0.17	0.14	0.22	1.19
County road (urban) × Rhône	[–]	[–]	[–]	1.00
County road (rural) × Ain	−0.64	0.12	<0.0001	0.53
County road (rural) × Ardèche	−0.46	0.15	0.002	0.63
County road (rural) × Drôme	−0.60	0.13	<0.0001	0.55
County road (rural) × Isère	−0.32	0.12	0.005	0.72
County road (rural) × Loire	−0.35	0.14	0.01	0.70
County road (rural) × Savoie	−0.78	0.14	<0.0001	0.46
County road (rural) × Haute-Savoie	−0.55	0.13	<0.0001	0.58
County road (rural) × Rhône	[–]	[–]	[–]	1.00
Over-dispersion	1.10	–	–	

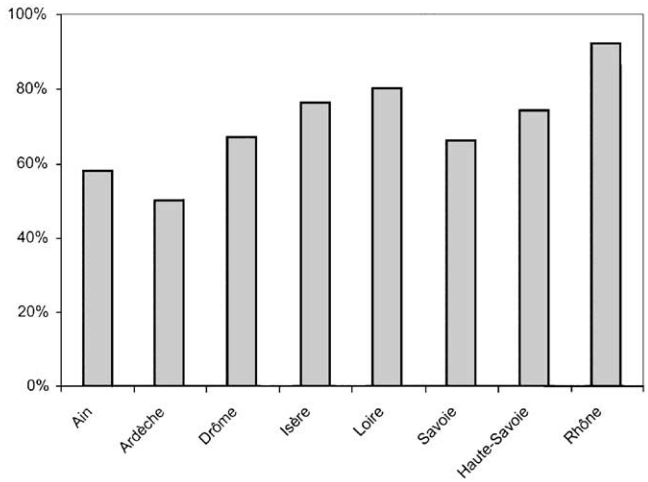

Fig. 3. Urbanisation rates.

means that it is also similar across road types. Severity is slightly increasing in all eight counties, on all road types (OR = 1.02 between any two consecutive years). The interaction between county and road type means that one can not compare globally the severity of the counties; this can only be done within each road type. Another way of saying it is that the ranking of counties according to their severity varies with the road type considered, and vice-versa. Results, including OR, are given in Table 5. We consider that the Odds Ratio (OR) estimates are a good approximation of RR estimates, as the average probability of a fatal crash for a given injury crash is small (0.074). Within each road type, one reads the OR of every county compared to the reference county (Rhône). For instance, severity on Drôme motorways is 1.9 times higher than on Rhône motorways, and severity on Drôme urban county roads is 0.53 times less than on Rhône similar roads.

5.3. *Traffic safety and county characteristics (driving and socio-economic data)*

In the analysis of incidence of injury accidents, the forward selection (after adjusting for road type and calendar trend) leads to the model including the following variables: proportion of people aged 35–64 (RR = 1.0), proportion of single persons (RR = 0.98), number of new driver licences—scaled by the number of vehicles—(RR = 1.43), number of accompanied learner drivers—scaled by the number of vehicles—(RR = 0.42).

In the analysis of severity of injury accidents, the forward selection (after adjusting for road type and calendar trend) leads to the model including the following variables: proportion of people aged $<$15 (RR = 1.15), number of suicides—scaled by the number of inhabitants—(RR $>$ 1000), number of new driver licences—scaled by the number of vehicles—(RR = 0.84), number of accompanied learner drivers—scaled by the number of vehicles—(RR = 2.28).

However, these results are likely to be unreliable. In both analyses, we encountered some odd behaviour in the forward selection of the model and sometimes obtained weird results (cf. estimate associated with suicides). We did not investigate this; the reason why is given in the discussion below.

6. Discussion and conclusion

The crash data come from police reports, and it is known that these data suffer from under-reporting. In this study, we compare ratios between counties and between road types. Under-reporting might biased the results of the analysis presented here if the under-reporting differs between counties or between road types. However, in the Rhone county, a study comparing Police reports and data from the road accident trauma registry of the Rhone county has found no differences in under-reporting between road types (Laumon and Martin, 2002). We do not foresee any reason why there would be difference in under-reporting between counties, but this can not be completely ruled out.

We have accounted for over-dispersion all through the modelling. As already mentioned, over-dispersion is common in crash data. Over-dispersion is also a way to account for unexplained systematic variation (Fridstrøm et al., 1995), and we are unfortunately in this case. First, we do not have a large set of independent variables available as possible risk factors. Secondly, we work on aggregated data, and it is known that such aggregated data suffer from loss of information (e.g. drawbacks of group exposure versus individual exposure, Estève et al., 1994) compared to studies based on individual data.

Over-dispersion can also hide mis-specification of the model. This is why it is necessary to perform checks of fit of the model.

Both analyses of incidence and severity accounting for different infrastructure yield to the same models, in the sense of including the same covariates. These variables are county, road types, calendar time assuming a linear trend, and the interaction between county and road type. Not only do differences in traffic safety among counties largely remain after adjusting on differences in road type distribution, but the differences are road-type dependent. To say it in another way, the ranking of counties in term of incidence or severity will vary depending on the road type considered.

However, for each index, time trend is the same across counties and across road types (incidence decreases, and severity increases slightly).

As already mentioned, county roads are managed by the counties, so that more heterogeneity in safety indexes (across counties) would be expected on this type of network. The results do not support this hypothesis: heterogeneity in incidence levels is highest on motorways; heterogeneity in severity levels seems similar on the different road types.

We had planned to analyse observed differences of traffic safety by first adjusting for road type, and then by including characteristics of the counties (driving and socio-economic data) to explore which of these could account for the remaining differences i.e. be a risk factor.

However, it appeared that there is interaction between county and road type. This means that characteristics of the counties which are defined at the level of the counties are not appropriate to account for remaining differences between the combination of counties and road type. This is the reason why we did not investigate the odd behaviour encountered in the modelling of these driving and socio-economic characteristics.

To explain differences in traffic safety among these county $\times$ road type combinations, we should look for driving and socio-economic data at this sub-level. For instance, it is empirically known that trips on Drôme and Loire motorways are of different type: on Drôme motorway, there is much traffic to the South for week-ends and vacations, whereas on Loire motorway, there are rather daily commuters between Saint-Etienne and Lyon (both being the largest city of their own county). This could lead to different driving behaviours and, hence, to different incidence and severity of crashes.

The main result of this study is that there is interaction between county and road type, meaning that differences in incidence and severity between counties depend on road type, and vice-versa, differences in incidence and severity between road types depend on county. To go further in the understanding of this, one needs to collect and explore socio-economic characteristics of these different road types within each county.

References

Andreassen, D., 1991. Population and registered vehicle data versus road deaths. Accident Anal. Prev. 23 (5), 343–351.

Aitkin, M., Anderson, D., Francis, B., Hinde, J., 1989. Statistical Modeling in GLIM. Oxford Science Publications, Oxford.

Baker, S.P., Whitfield, R.A., O'Neill, B., 1987. Geographic variations in mortality from motor vehicle crashes. The N. Engl. J. Med. 316 (22), 1384–1387.

Breslow, N.E., Day, N.E., 1987. Statistical Methods in Cancer Research: The Design and Analysis of Cohort Studies, Vol. II. IARC Scientific Publications, No. 82. International Agency for Research on Cancer, Lyon.

Brühning, E., Alevisos, E., 1991. IRTAD International Road Traffic and Accident Database. Zeitschrift für Vekhersicherheit, BAST.

Chipman, M.L., McGregor, C.G., Smiley, A.M., Lee-Gosselin, M., 1992. Time versus distance as measures of exposure in driving surveys. Accident Anal. Prev. 24 (6), 679–684.

Estève, J., Benhamou, E., Raymond, L., 1994. Statistical Methods in Cancer Research: Descriptive Epidemiology, Vol. IV. IARC Scientific Publications, No. 128. International Agency for Research on Cancer, Lyon.

Fridstrøm, L., Ingebrigtsen, S., 1991. An aggregate accident model based on pooled, regional time-series data. Accident Anal. Prev. 23 (5), 363–378.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., Krogsgård Thomsen, L., 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accidents counts. Accident Anal. Prev. 27 (1), 1–20.

Gourieroux, C., Montfort, A., 1984a. Pseudo maximum likelihood methods: theory. Econometrica 52, 681–700.

Gourieroux, C., Montfort, A., 1984b. Pseudo maximum likelihood methods: application to Poisson models. Econometrica 52, 701–720.

Greenwood, M., Yule, G., 1920. An inquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. J. R. Stat. Soc., Ser. A 83, 255–279.

Hemenway, D., 1998. The decomposition method. Epidemiology l9 (4), 369–370.

Joly, M.F., Joly, P., Bergeron, J., Desjardins, D., Ekoe, J.M., Ghadirian, P., Gravel, S., Hamet, P., Laberge-Nadeau, C., 1991. L'exposition au risque d'accident de la route, un paramètre épidémiologique fondamental et difficile à mesurer. Revue d'Epidémiologie et de Santé Publique 39, 307–313.

Kerrich, J., 1951. Accident Statistics and the Concept of Accident-Proneness. Part II. The Mathematical Background. Biometrics 7, 391–432.

King, G., 1989. Variance specification in event count models: from restrictive assumptions to a generalized estimator. Am. J. Pol. Sci. 33 (3), 762–784.

Lassarre, S., Hoyau, P.A., 1995. La hausse de la gravité en France est-elle due à un effet de structure? Les Cahiers de l'Observatoire, 10, 57–75, La Documentation Francaise, Paris.

Laumon, B., Martin, J.L., 2002. Analyse des biais dans la connaissance épidémiologique des accidents de la route en France. Revue d'Epidémiologie et de Santé Publique, in press.

Laumon, B., Martin, J.L., Collet, P., Chiron, M., Ndiaye, A., Verney, M.P., Vergnes, I., 1997. A French road accident trauma registry: first results. In: Proceedings of the 41st AAAM's Annual Conference, Vol. 127–137. AAAM Eds, Des Plaines.

Li, G., Baker, S.P., Langlois, J.A., Kelen, G.D., 1998. Are female drivers safer? An application of the decomposition method. Epidemiology 9 (4), 379–384.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd Edition. Chapman & Hall, London.

Mannering, F.L., 1993. Male/female driver characteristics and accident risk: some new evidence. Accident Anal. Prev. 25 (1), 77–84.

Mercer, G.W., 1987. Influences on passenger vehicle casualty accident frequency and severity: unemployment, driver gender, driver age, drinking driving and restraint device use. Accident Anal. Prev. 19 (3), 231–236.

Shapiro, S., Wilk, M., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.