Prediction of responsibility for drivers and riders involved in injury road crashes

Cédric Garcia, Vivian Viallon, Liacine Bouaoun, Jean-Louis Martin

Prediction of responsibility for drivers involved in injury road crashes

Cédric Garcia[a], Vivian Viallon[a,b], Liacine Bouaoun[a,b], Jean-Louis Martin[a]

[a] Univ Lyon, Université Lyon 1, IFSTTAR, UMRESTTE UMR_T 9405, F-69675 Lyon, France

[b] International Agency for Research on Cancer, Lyon, France

Abstract

Responsibility analysis allows the evaluation of crash risk factors from crash data only, but requires a reliable responsibility assessment. The aim of the present study is to predict expert responsibility attribution (considered as a gold-standard) from explanatory variables available in crash data routinely recorded by the police, according to a data-driven process with explicit rules. Driver responsibility was assessed by experts in the light of all information contained in police reports for a sample of about 5,000 injury crashes that occurred in France in 2011. Three statistical methods were implemented to predict expert responsibility attribution: logistic regression with L1 penalty, random forests, and boosting. Potential predictors of expert attribution referred to inappropriate driver actions and to external conditions at the time of the crash. Logistic regression was finally chosen to construct a score to assess responsibility for drivers and riders in crashes involving one or more motor vehicles, or involving a cyclist or pedestrian. Cross-validation showed that our tool can predict expert responsibility assessments on new data sets. In addition, responsibility analyses performed using either the expert responsibility or our predicted responsibility return similar odds ratios. Our scoring process can then be used to reliably assess responsibility based on national police report databases, provided that they include the information needed to construct the score.

Keywords

road crash; responsibility analysis; scoring tool; statistical learning; crash risk factor

Accepted for publication in Journal of Safety Research on July 2, 2019 https://doi.org/10.1016/j.jsr.2019.07.001

1. Introduction

Injury road crashes are a rare event for drivers (Blaizot, Papon, Haddak, & Amoros, 2013; Bouaoun, Haddak, & Amoros, 2015). The study design preferred in epidemiology to assess risk factors is therefore case-control analysis, with casualties as cases and drivers with no crashes over a given period as controls. For temporary risk factors such as driving under the influence or using a phone at the wheel, the comparison is between casualties (cases) and other users on the road during the period in which the crash happened (controls) (Sagberg, 2001). For enduring risk factors such as age, gender or health status, comparison is between casualties (cases) and drivers (controls) with the same level of exposure (Brubacher, Chan, & Asbridge, 2012). However, it is difficult in these studies to use road checks to assess risk factors such as using a phone at the wheel or being distracted. Indeed, the cooperation of some drivers, when questioned in a check, for example, for driving under the influence of alcohol or drugs, will vary depending on whether they consider themselves at fault or not. It is actually very difficult to estimate crash risk exposure in a population of drivers.

To overcome the problem of not having an appropriate control group, methods using the concept of crash responsibility (or at-fault) have been developed (Haight, 1973; Lyles, Stamatiadis, & Lighthizer, 1991; Stamatiadis & Deacon, 1997; Thorpe, 1964). The technique estimates the risk of being responsible for a crash for each road user involved, simply relying on road crash databases. The approach hypothesizes that the set of non-responsible drivers is a representative sample of drivers as a whole (af Wåhlberg & Dorn, 2007; Brubacher, Chan, & Asbridge, 2014). The hypothesis is based on the fact that those responsible for the crash do not deliberately choose which other drivers are going to be involved, and that consequently all non-responsible drivers have the same risk of being involved in a crash (Chandraratna & Stamatiadis, 2009). Following Davis and Gao (Davis & Gao, 1995), non responsible victims are assumed to be selected by the responsible (at-fault) driver randomly from the pool of available drivers (Cooper, Meckle, & Andersen, 2010), with the probability that the non-responsible victim is the member of a given subgroup being directly proportional to that subgroup's exposure at the accident site.

The validity of this method presupposes that it is possible to assess precisely the degree of responsibility of each of those involved in the crash. This means having an exact description of the circumstances. Responsibility here is not defined in any legal sense: a road user is deemed "responsible" if he contributes to or triggers the crash, typically by a faulty maneuver (driving the wrong way on a one-way road, running a red light, etc.) or failure (to brake in time, to switch on headlights at night or in a tunnel, etc.). It is therefore essential that the definition of responsibility should be based directly on such behavioral factors and not on their underlying causes such as inexperience, alcohol consumption, using the phone at the wheel, etc. Otherwise, the impact of these factors on the risk of being responsible for a road crash would be greatly overestimated.

In France, the police draw up a free-text report for any injury road crash they have been called to. Some of this information is routinely computerized in a police record database (PRDB), recording a variety of details such as place of crash, vehicles involved, road users involved and infringements. In particular, the police, in drawing up the report, detail the responsibility of each party, as they see it. These police attributions of responsibility can be used for the purposes of responsibility analysis, but the criteria are not clearly laid down and the validity of the attribution is not guaranteed. The police are liable to have an unduly legalistic attitude, which may not fit with the definition of responsibility given above.

Ideally, reliable attribution of responsibility requires analysis by a trained expert working from a finely detailed description of the crash. This was done in the VOIESUR project (*Véhicule Occupant Infrastructure Etudes de la Sécurité des Usagers de la Route*: Vehicle Occupant Infrastructure Studies of Road User Safety) in 2011, providing a reliable criterion of responsibility (meaning contributing to a crash) that was as objective as possible: i.e., fact-based. Responsibility as defined in the VOIESUR database is thus considered optimally reliable for the purposes of responsibility analysis.

The aim of the present study was to estimate responsibility according to a data-driven process with explicit rules. With this aim, several methods of statistical learning were compared, with cross-validation to avoid overfitting, to predict experts' responsibility attributions (considered as gold-standard) from data routinely recorded by the police.

2. Material and methods

2.1. Data

As part of the VOIESUR project, police reports drawn up in 2011 (by the two forces operating in France: *Police* and *Gendarmerie*) were digitized and centralized by the TransPV organization on behalf of insurance companies. The data source providers were contacted in case of important missing data such as crash scene diagrams, vehicle photographs and injury assessments.

The database recorded all fatal crashes and one-twentieth of injury crashes that occurred in 2011: 7,846 crashes in metropolitan France for which information was available about crash configuration and location, vehicle photographs, each road-user's actions before the crash (including any infringements), collisions and identified relevant conflicts. Textual information, often written by police officers, was also used to shed light on circumstances. In all, more than 300 variables describing the crashes were available.

A team of experts from the VOIESUR project had access to this information, in order to determine road-user responsibilities. The responsibility variable given by experts was graded as: 1, user completely responsible; 2, user fairly responsible, the contribution of the user to the crash possibly had some external factor out of its control; 3, shared responsibility; 4, user fairly non-responsible, the user could maybe have avoided the crash; and 5, user totally non-responsible.

2.2. Crash configurations

To estimate the degree of driver responsibility, it is essential to take account of other relevant road users' behavior (drivers, cyclists or pedestrians). This information varies according to the type of road user: speeding is not relevant for cyclists, and changing lanes means nothing in the case of pedestrians.

For simplicity, we only considered the most frequent crash configurations; those involving 2 or more cyclists, or 1 cyclist and 1 pedestrian were not considered:

- Configuration 1: crash involving only motor vehicles, 2 or more;
- Configuration 2: crash involving a motor vehicle and a pedestrian or a cyclist;
- Configuration 3: crash involving only 1 motor vehicle.

2.3. Definition of outcome and coding of study variables

The study objective was to predict whether a driver involved in a crash could be considered as being responsible for it. To this end, we first opted for binary coding of expert attributions. More precisely, drivers were deemed responsible if the expert grade was 1 or 2. This choice to include "fairly responsible" (grade 2) cases was based on the following: although crashes often occur due to a combination of factors, eliminating any one of them would usually lead to the crash not happening. In other words, it was considered that the crash would not have happened if the road user had not done whatever it was that led the expert to deem him or her "fairly responsible". The non-responsible group comprised drivers with responsibility graded 4 or 5.

Grade 3 responsibility was assessed for 6% of drivers. These drivers were excluded from analysis, because we have considered more efficient to base the learning process on cases where responsibility was clearly attributed mostly to one driver. Binary expert attribution was noted as "Y", such that Y=0 for non-responsible drivers and Y=1 for responsible drivers.

Accepted for publication in Journal of Safety Research on July 2, 2019 https://doi.org/10.1016/j.jsr.2019.07.001

For explanatory variables, we focused on those found in the PRDB database referring to inappropriate actions that could have led to the crash. We did not take account of the possible causes of such inappropriate actions, which are the risk factors generally studied in responsibility analyses (alcohol, cannabis, telephone at the wheel, etc.), as the aim was to achieve a final prediction of expert attribution that would be independent of the factors underlying inappropriate actions, as explained above. As potential predictors of expert attribution we therefore considered PRDB variables referring to actions, such as driving the wrong way on a one-way road, speeding, failure to give way, making a half-turn or overtaking (on the right or on the left), etc. For a given road user *i,* such variables were formalized as $(Z_{i1},\dots,Z_{iq})$.

We also included as potential predictors some variables referring to external conditions at the time of the crash: weather, road surface, etc. The reason to include them was we believe some can alleviate responsibility of a driver, such as rain which reduces visibility, and then reduce the responsibility of all drivers involved in the corresponding crash. On the other hand, road type such as two-way road indicates that deviating on the left increases the chances of a head-on collision, giving better prediction on responsibility. These external conditions variables were formalized as $(W_{i1},\dots,W_{ip})$. Their values are obviously the same for all road users involved in a given crash. All these variables, $(Z_{i1},\dots,Z_{iq})$ and $(W_{i1},\dots,W_{ip})$, are numerical, whether continuous or binary. The set of explanatory variables could vary according to the type of crash. In the end the variables chosen for crash configuration 1 and 3 are shown in Table 1. For configuration 2, variables “Pedestrian masked, playing or running” and “Pedestrian on pedestrian crossing” were also considered.

Accepted for publication in Journal of Safety Research on July 2, 2019 https://doi.org/10.1016/j.jsr.2019.07.001

Table 1 : List of all explanatory variables included for configuration 1 and configuration 3 crashes

Notation	Type of variable	Group of categorical variable	Variable
W	External conditions	Number of lanes	One-way road
			2-lane 2-way road
			3-lane 2-way road
			4-lane 2-way road or dual carriageway
			Road with separated lanes
		Type of road	Main road
			Urban road
			B road
		Particular layout on the road	Road bridge, tunnel or subway
			Ramp
			Crossroads
		Special lane	Presence of special lane (cyclist lane, road toll, etc.)
		Presence of intersection	X intersection
			T or Y intersection
			Roundabout
			Other
		Light	Night without road lighting
			Night with equipped crossroads
		Weather	Heavy rain
			Fog, snow or storm
			Bad weather
		State of road	Steep slope
			Curved road
			On central reservation
			Water, mud, ice or oil on the road
		Crash localization	Crash not on road
		Special event	Party the day of crash or the day before
		Spacing	Width of the road (numeric)
Z	Vehicle impact	Localization of the vehicle impact	Rear
			Frontal
			Left side
			Right side (= passenger side)*
		Mobility of the hit obstacle	Mobile obstacle
			Fixed obstacle
	Driver actions	Mobility of the vehicle	Stationary vehicle
			Speeding
		Maneuver of the vehicle	Avoidance maneuver
			Left turn (= into traffic)*
			Overtaking on the left (= normal overtaking)*
			Vehicle deviating left (= into traffic)*
			Turning left or right
			Insertion of vehicle into traffic
			Overtaking a vehicle
			Between 2 lanes, half-turn or reversing
			Changing lanes
			Vehicle deviating
			Vehicle change of direction without indication
		Dangerous behavior	Intended imprudence, dangerous overtake or no-way street
			Forbidden road
			Failure to give way
		Count of faults	Number of faults (numeric)

*Driving is on the right in France.

Accepted for publication in Journal of Safety Research on July 2, 2019 https://doi.org/10.1016/j.jsr.2019.07.001

The last variable pointing at faults are categorized as follows:

- Vehicles:
- Unannounced change of direction
- Intended imprudence, dangerous overtake or no-way street
- Forbidden road
- Failure to give way
- Between two lanes, U-turn or reverse maneuver
- Speed higher than limit
- Stopping or parking wrongly
- Vehicle's lights not turned on in obscurity or night
- Tailgating
- Speed too slow for the road category
- Pedestrians:
- Pedestrian on pavement but not on crosswalk
- Pedestrian running or hidden

2.4. Construction of a predictive model

To discriminate responsible from non-responsible drivers, we applied three different statistical methods for each crash configuration: logistic regression with L1 penalty, random forests, and boosting.

Logistic regression and LASSO penalization

For a road-user *i* in the VOIESUR database, let $Y_i \in \{0,1\}$ be the binary variable of responsibility attributed on initial expert coding, and $X_i \in R^m$, for $m \geq 1$, be the vector of the accepted predictors; this vector will be deduced from PRDB variables $(Z_{i1},\ldots,Z_{iq})$ and $(W_{i1},\ldots,W_{ip})$ which are introduced according to the type of crash. See below for more detail. The logistic regression presupposes parameters $\alpha \in R$ and $\beta \in R^m$ such that:

$$logit\big(P(Y_i = 1|X_i)\big) = log\left(\frac{P(Y_i = 1|X_i)}{1 - P(Y_i = 1|X_i)}\right) = \alpha + X_i^T\beta$$

which can be expressed as:

$$P(Y_i = 1\ |X_i) = \frac{\exp(\alpha + X_i^T\beta)}{1 + \exp(\alpha + X_i^T\beta)} \qquad (1)$$

As the value of m is high (54 explanatory variables), the values of parameters α and β are estimated by maximizing L1-norm penalized likelihood, to select the most relevant predictors and improve predictive performance. With $\mathrm{L}(\alpha,\beta)$ being the log-likelihood of the logistic regression model, the LASSO (least absolute shrinkage and selection operator) logistic regression estimates parameters by values maximizing the penalized criterion $(\mathrm{L}(\alpha,\beta) - \lambda||\beta||_1)$, by choosing the appropriate regularization parameter λ. In the present case, the number of predictors is high, but not compared to the number of observations. We therefore adopted a two-stage version, following the ideas of the LASSO-OLS Hybrid (Efron, Hastie, Johnstone, & Tibshirani, 2004). In the first stage, the LASSO logistic regression was used to select relevant predictors, the ones not removed by the penalized likelihood. In the second stage, logistic regression with maximum non-penalized likelihood was used to re-estimate the parameters of the model corresponding to the predictors kept in the first stage (this is the “standard” OLS-Hybrid strategy suggested by Efron et al. 2004). We therefore chose the λ value from the first stage that minimized the Akaike information criterion (AIC) calculated on the basis of the values obtained in stage 2, with the non-penalized likelihood. Every categorical variable with K classes was dichotomized into K-1 dummy variables, which were then independently selected when applying the LASSO penalty.

Having established this general principle, we can specify how the predictor vector X_i is constructed from the variables $(Z_{i1},\ldots,Z_{iq})$ and $(W_{i1},\ldots,W_{ip})$. The construction depends on the type of crash. In type-3 crashes (involving a single vehicle), vector X_i is simply equal to $(Z_{i1},\ldots,Z_{iq},W_{i1},\ldots,W_{ip})$. Type-1 and -2 crashes involve more than 1 road user, and responsibility depends not only on each user's own actions but also on those of the others. Take the example of a 2-vehicle collision involving drivers A_1 and A_2. The fact that A_1 entered a crossroads against a red light increases his or her likelihood of being deemed responsible by the expert, and reduces the likelihood for A_2. Preliminary results (data not shown) confirmed that the increased risk for A_1 is of similar magnitude on a logit probability scale to the decrease for A_2. In such cases, we therefore replaced each variable Z_{1j} for A_1 by $S_{1j} = Z_{1j} - Z_{2j}$ and each variable Z_{2j} for A_2 by $S_{2j} = Z_{2j} - Z_{1j}$. In crashes involving more than 2 road users, a similar principle was applied. Let denote by A_1, …, A_l the l road users involved in a given crash and Z_{ij} the j^{th} covariable of inappropriate action by A_i. Then Z_{ij} is replaced by $S_{ij} = Z_{ij} - \max_{k\neq i} Z_{kj}$, with the maximum calculated for the whole set of *l-1* antagonists of A_i. Note that, in the case where *l=2*, this rule comes down to the same thing as that described above for crashes involving 2 road users.

Now, let $\tilde{X}_i = (S_{i1},\ldots,S_{iq},W_{ip},\ldots,W_{ip})$. With the above calculation rule, the impact of certain predictors of expert attribution included in $\tilde{X}_i$ may vary depending whether the crash involved 2 or ≥3 road users. In type-1 crashes, let T_i be the binary variable indicating whether the crash involved 2 ($T_i=0$) or ≥3 road users ($T_i=1$). All type-1 crashes will be considered using a model including possible interactions between the components of $\tilde{X}_i$ and T_i. More precisely, the model used is:

$$\text{logit}[P(Y_i = 1\,|\tilde{X}_i, T_i)] = \alpha + \tilde{X}_i^T\tilde{\beta} + T_i\tilde{X}_i^T\tilde{\gamma} \qquad (2)$$

which can be reformulated as

$$\text{logit}[P(Y_i = 1\,|X_i)] = \alpha + X_i^T\beta \qquad (3)$$

where $X_i^T = (\tilde{X}_i^T, 0_{m/2}^T) \in R^m$ if 2 vehicles are involved, and $X_i^T = (\tilde{X}_i^T, \tilde{X}_i^T) \in R^m$ if ≥3 vehicles are involved, with m = 2(p + q), 0_r the null vector of R^r, and $\beta = (\tilde{\beta}, \tilde{\gamma})$. The LASSO L1 penalty favors null status for components of vector β (and thus of vectors $\tilde{\beta}$ and $\tilde{\gamma}$), and thus absence of interaction. Nevertheless, it allows the most relevant non-null components to be identified, especially in vector $\tilde{\gamma}$, which contains the interaction terms. If we note $R = \alpha + X_i^T\beta$, then equation (1) implies that if R>0, $P(Y_i = 1\,|X_i) > P(Y_i = 0\,|X_i)$. The corresponding driver is more likely responsible than non-responsible according to our prediction model, hence declared responsible.

This is the optimal choice to minimize the number of misclassified, as the average percentage of responsible drivers is around 50% in type 1 and type 2 crashes.

A similar principle was applied to drivers in type-2 crashes, where the influence of predictors in $\tilde{X}_i$ can vary according to the type of third party (pedestrian or cyclist): for any driver *i* involved in a type-2 crash, we constructed the predictor vector X_i as: $X_i^T = (\tilde{X}_i^T, 0_{m/2}^T) \in R^m$ if a pedestrian is involved in the crash, and $X_i^T = (\tilde{X}_i^T, \tilde{X}_i^T) \in R^m$ if a cyclist is involved, with, again, *m = 2(p + q)* and 0_r the null vector of R^r. Note that observations for pedestrians and cyclists are used only to describe their actions with respect to drivers whose responsibility is being estimated.

The score R is then calculated as:

$$R = \alpha + X_i^T\beta$$

If R>0, the driver *i* is considered responsible, and otherwise non-responsible.

Decision tree methods /Machine learning algorithms

As mentioned above, as well as logistic regression with LASSO penalty, we also examined decision-tree classifications. These methods seem well suited to identifying responsible drivers for each type of crash, since expert attribution is based on an implicit decision tree: if the driver commits a “serious” fault such as running a red light, this is usually enough for the expert to attribute responsibility. If no serious faults were committed, the driver’s other actions need to be examined in detail to determine responsibility, going further down the decision-tree. We tested two classical decision-tree approaches: boosting and random forests (Hastie, Tibshirani, & Friedman, 2011).

The random forests is a variant of the bagging method. Consider our training data $\mathrm{Z} = \{(\mathrm{X}_1, \mathrm{Y}_1), \dots, (\mathrm{X}_n, \mathrm{Y}_n)\}$, suppose we fit a model from this data, obtaining the prediction $\hat{\mathrm{f}}(\mathrm{X})$ at input X. We construct a bootstrap sample $\mathrm{Z}^{*\mathrm{b}}, \mathrm{b} = 1, 2, \dots, \mathrm{B}$, we fit the decision tree model for every bootstrap sample, giving prediction $\mathrm{f}^{*\mathrm{b}}(\mathrm{X})$. Over the B samples, the bagging estimate is the class with the highest proportion of trees predicting it. Random forests are decision trees applied on bagging model, but for each node of a tree, Random forests randomly selects a fixed number of explanatory variables to use in order to split the node. This choice of random variables for each bootstrap sample allows reducing the variance of the bagging estimate.

Boosting method consists in creating weak classification trees to repeatedly modified version of the data. The data changes at each boosting step consist of putting weights $(\omega_1, \dots, \omega_n)$ to each of the training observations $(\mathrm{X}_1, \mathrm{Y}_1), \dots, (\mathrm{X}_n, \mathrm{Y}_n)$, starting from equal weights for the first classifier. For a chosen M number of weak classification trees, each iteration increases the weights of misclassified observations from the last classifier. For an input X, The final classifier G(X) will be the weighted sum of all $\mathrm{G}_1(\mathrm{X}), \dots, \mathrm{G}_\mathrm{M}(\mathrm{X})$ classifiers, with $(\alpha_1, \dots, \alpha_\mathrm{M})$ weights calculated as $\alpha_\mathrm{m} = \log\left(\frac{1}{\mathrm{err}_\mathrm{m}} - 1\right)$, err_m being the error rate of classifier G_m on the training data. By scaling these weights to make their sum equal to 1, if $\mathrm{G}(\mathrm{X}) > 0.5$ then the driver will be predicted responsible, else it will be predicted as non-responsible.

Validation and comparison of methods

The accuracy of the models’ predictions applied to new data sets was assessed by implementing a K-fold cross-validation method for the logistic regression and for the boosting. The random forest method does not require cross-validation, as it makes an out-of-bag estimate of prediction error.

Results on the three models were compared in two ways:

- First, on the following five criteria, with prediction noted as $\hat{Y}$:
  accuracy, defined as the percentage of observations in which expert attribution matched prediction ($Acc = P(\hat{Y} = Y)$); sensitivity, defined by $S_n = P\left(\hat{Y} = 1 \middle| Y = 1\right)$; specificity, defined by $S_p = P\left(\hat{Y} = 0 \middle| Y = 0\right)$; area under the ROC curve (AUC), defined by the area of the curve plotting the sensitivity against 1 – specificity for varying risk-thresholds; and Cohen kappa, an inter-rater agreement measure for contingency tables, with 0 corresponding to independence of judgments and 1 corresponding to perfect agreement;
- Second, by comparing ORs for several risk factors, derived from a responsibility analysis using either expert attribution or predicted responsibility.

All statistical methods were implemented using R statistical software. Logistic regression with LASSO selection was implemented using the glmnet R package. Random forest and boosting algorithms were implemented using the randomForest and gbm R packages, respectively.

3. Results

3.1. Sample size

Table 2 presents number of crashes, number of drivers (or riders) and percentage of responsible drivers (i.e., *Y=1*) per crash configuration. As expected, percentage responsible drivers varied according to crash configuration, at around 50% for types 1 and 2 (involving ≥2 road users) and almost 100% for type 3 (involving a single motor vehicle).

Table 2: Descriptive statistics of the three crash configurations.

	Configuration		
	Type 1: at least two motor vehicles	Type 2: motor vehicle and a pedestrian or cyclist	Type 3: single vehicle
Number of crashes	3,583	1,262	1,961
Number of drivers involved	7,597	1,262	1,961
% responsible drivers or riders	47.1%	51.4%	98.1%

In the following subsections, we successively present classification algorithm results for each of the 3 types of crash.

3.2. Crash involving at least two motor vehicles (configuration 1)

We first assessed the predictive capacities of the algorithms: i.e., their ability to correctly predict the expert attributions. Table 3 presents results in terms of accuracy, sensitivity, specificity, AUC and Cohen kappa for the LASSO logistic regression, random forests and boosting. The LASSO logistic regression showed similar values for sensitivity and specificity; sensitivity was higher but specificity lower than for decision-tree-based approaches. Even so, performance was largely similar for all three methods on the global criteria of accuracy, AUC and kappa, indicating very good performance for all three, so that all three could be used for precise prediction of expert attributions of responsibility. Overall, we favor logistic regression because the derived score has an explicit form, unlike the other two.

Table 3: Results on various criteria, using cross-validation for logistic regression and boosting, and out-of-box estimate for random forests

	LASSO logistic regression	Random forests	Boosting
Accuracy	0.869 [0.862 ; 0.877]	0.864 [0.856 ; 0.871]	0.869 [0.862 ; 0.876]
Sensitivity	0.887 [0.877 ; 0.898]	0.812 [0.799 ; 0.825]	0.837 [0.825 ; 0.849]
Specificity	0.853 [0.843 ; 0.864]	0.909 [0.900 ; 0.918]	0.898 [0.889 ; 0.906]
AUC	0.936 [0.931 ; 0.942]	0.932 [0.927 ; 0.938]	0.936 [0.931 ; 0.940]
Cohen kappa	0.739 [0.724 ; 0.754]	0.725 [0.710 ; 0.741]	0.737 [0.722 ; 0.751]

Validation through responsibility analysis

An additional way to assess the validity of our predicted responsibility is to assess whether responsibility analysis conducted with either the expert responsibility or our predicted responsibility returns similar odds ratios for some risk factors of interest (typically not included among the predictors in our models). Table 4 shows ORs estimated for six selected risk factors using, in turn, expert attributions of responsibility and responsibility estimated by the LASSO strategy. We chose to perform this analysis with the six following risk factors: driving under the influence of alcohol (reference: <0.5g/l), age (by age group; reference: <20 years), gender (reference: male), socio-occupational class (reference: executive of manual worker), driver's license (reference: valid license) and type of road user (reference: driver).

Accepted for publication in Journal of Safety Research on July 2, 2019 https://doi.org/10.1016/j.jsr.2019.07.001

Table 4: Odds ratio of expert attribution of responsibility and LASSO predicted responsibility with 95% confidence intervals, for 6 risk factors

Variable	Category	Expert	LASSO
Blood alcohol concentration (g/L)	[0 ; 0.5[	Reference	Reference
	[0.5 ; 0.8[	3.19 [1.70 ; 6.00]	2.36 [1.29 ; 4.33]
	[0.8 ; 1.2[	6.92 [3.81 ; 12.59]	6.85 [3.70 ; 12.70]
	[1.2 ; 2.0[	14.60 [7.60 ; 28.03]	6.32 [3.86 ; 10.38]
	2.0 or more	13.45 [7.40 ; 24.48]	10.19 [5.83 ; 17.82]
	≥0.5 but unknown precise value	7.83 [3.03 ; 20.23]	2.91 [1.40 ; 6.05]
	Unknown value	1.40 [1.21 ; 1.62]	1.31 [1.13 ; 1.51]
Age	Less than 20 years	Reference	Reference
	[20 ; 40[	0.65 [0.53 ; 0.82]	0.64 [0.51 ; 0.80]
	[40 ; 60[	0.48 [0.38 ; 0.61]	0.49 [0.38 ; 0.61]
	60 years or more	0.73 [0.53 ; 1.00]	0.73 [0.53 ; 1.00]
Sex	Man	Reference	Reference
	Woman	0.88 [0.78 ; 0.98]	0.91 [0.81 ; 1.02]
Socio-professional class	Executive of manual worker	Reference	Reference
	Farmer, artisan or storekeeper	0.82 [0.65 ; 1.02]	0.85 [0.68 ; 1.06]
	Professional driver	1.34 [1.04 ;1.72]	1.15 [0.90 ; 1.48]
	Retired	1.39 [1.07 ; 1.81]	1.23 [0.95 ; 1.60]
	Unemployed or student	1.61 [1.33 ; 1.95]	1.55 [1.28 ; 1.88]
	Other	1.03 [0.92 ; 1.16]	1.10 [0.98 ; 1.24]
Driving License	Valid license	Reference	Reference
	Invalid or lack of driving license	2.85 [1.78 ; 4.55]	2.46 [1.57 ; 3.87]
Driver type	Car	Reference	Reference
	Motorcycle	0.90 [0.80 ; 1.02]	0.91 [0.80 ; 1.03]
	Truck, bus or other	0.45 [0.36 ; 0.57]	0.53 [0.43 ; 0.66]

Overall, Table 4 shows that using responsibility predicted by the LASSO logistic regression or expert attributions lead to similar estimated ORs, even if different conclusions regarding statistical significance may arise for ORs close to 1. ORs for blood alcohol level were similar for levels <1.2g/l, but for levels above 1.2g/l, the use of predicted responsibility led to lower ORs.

Accepted for publication in Journal of Safety Research on July 2, 2019 https://doi.org/10.1016/j.jsr.2019.07.001

Responsibility prediction score

Finally, we used the logistic regression model to build our responsibility score.

Table 5: Coefficient values to construct the responsibility score for crashes involving just 2 motor vehicles (column 3) or more than 2 (column 4)

<table>
<tr><th>Type of variable</th><th>Variable</th><th>Value of coefficient β for crash involving 2 vehicles</th><th>Value of coefficient β for crash involving ≥3 vehicles</th></tr>
<tr><td>Intercept</td><td>Intercept</td><td>$\alpha = 0.023$</td><td>$\alpha = 0.185$</td></tr>
<tr><td>External conditions</td><td>One-way road</td><td>0</td><td>−0.688</td></tr>
<tr><td></td><td>4-lane 2-way road or dual carriageway</td><td>0</td><td>−0.553</td></tr>
<tr><td></td><td>Main road</td><td>0</td><td>−0.361</td></tr>
<tr><td></td><td>Urban road</td><td>0</td><td>−0.466</td></tr>
<tr><td></td><td>Bridge, tunnel or subway</td><td>0</td><td>−0.466</td></tr>
<tr><td></td><td>Heavy rain</td><td>0</td><td>0.629</td></tr>
<tr><td></td><td>Fog, snow or storm</td><td>0</td><td>0.979</td></tr>
<tr><td>Vehicle impact</td><td>Rear</td><td>−1.353</td><td>−0.730</td></tr>
<tr><td></td><td>Frontal</td><td>0</td><td>1.194</td></tr>
<tr><td></td><td>Right side (= passenger side)*</td><td colspan="2">0.062</td></tr>
<tr><td></td><td>With mobile obstacle</td><td>0.198</td><td>0.430</td></tr>
<tr><td></td><td>With fixed obstacle</td><td colspan="2">0.524</td></tr>
<tr><td>Driver actions</td><td>Stationary vehicle</td><td colspan="2">−0.511</td></tr>
<tr><td></td><td>Speeding</td><td>0</td><td>0.245</td></tr>
<tr><td></td><td>Avoidance maneuver</td><td>0</td><td>0.576</td></tr>
<tr><td></td><td>Left turn (= into traffic)*</td><td colspan="2">0.143</td></tr>
<tr><td></td><td>Overtaking on the left (= normal overtaking)*</td><td colspan="2">0.215</td></tr>
<tr><td></td><td>Vehicle deviating left (= into traffic)*</td><td colspan="2">1.098</td></tr>
<tr><td></td><td>Turning left or right</td><td colspan="2">1.133</td></tr>
<tr><td></td><td>Insertion of vehicle into traffic</td><td colspan="2">1.469</td></tr>
<tr><td></td><td>Overtaking a vehicle</td><td colspan="2">1.533</td></tr>
<tr><td></td><td>Between 2 lanes, half-turn or reversing</td><td colspan="2">1.636</td></tr>
<tr><td></td><td>Changing lanes</td><td colspan="2">2.032</td></tr>
<tr><td></td><td>Vehicle deviating</td><td colspan="2">2.064</td></tr>
<tr><td></td><td>Number of faults</td><td>1.897</td><td>2.217</td></tr>
</table>

*Driving is on the right in France.

The variable "Number of faults" is quantitative; all the others are qualitative, with values of $-1, 0,$ or 1. Indeed as defined in paragraph 2.3, in the case of two drivers A_1 and A_2, the value of S for A_1 is $S_{A_1} = Z_{A_1} - Z_{A_2}$.

For some variables, such as "Stationary vehicle", coefficients were equal whatever the number of motor vehicles involved. Others had a coefficient that depends on the number of vehicles: in particular, speeding significantly increased the risk of being responsible only for crashes involving at least three vehicles (Table 5). External conditions (weather and road conditions) were associated with the risk of being responsible for crashes involving at least three vehicles but not for crashes involving only two vehicles. It is also noteworthy that frontal impact has no influence for two-vehicle crashes, but has a strong influence in case of three vehicles or more, which may be explained by crashes occurring in traffic jams. As expected, we observe that committing a fault, making a direction-changing maneuver, hitting an opponent car on its rear (S=-1 for the corresponding driver) are the factors that increase the most the responsibility score, while they greatly reduce the responsibility score for a driver if the opponent car driver makes these errors.

Some variables are linked by construction: thus, the score for a driver “turning left or right” increases by 1.13, and again by 0.143 if it is actually turning left (i.e., into the traffic). In particular, deviating on left gives the highest contribution to responsibility score, as it often results in head on collision.

3.3. Crashes between a motor vehicle and a pedestrian or cyclist (configuration 2)

The predictive models used in the present work are above all based on variables indicating the driver behavior. They are not well-adapted for predicting responsibility in cyclists, and even less pedestrians. For example, some important information was missing for pedestrians and cyclists, such as crossing a red light or failure to give way to a tram. We therefore restricted our prediction objective to motor vehicle drivers (or riders).

1,262 crashes were recorded in that configuration, 946 of which involved a driver and a pedestrian and 316 a driver and a cyclist.

The performance of the three methods is presented in Table 6. All three methods had poorer performance in configuration 2 than 1. Accuracy for logistic regression was about 0.77 in configuration 2 and 0.87 in configuration 1. Cohen kappa showed moderate agreement between prediction and expert attribution. In configuration 2, the three methods again performed similarly, and we therefore decided to keep the LASSO logistic regression model for estimating the prediction score, due to its simpler form.

Table 6: Results on various criteria by cross-validation or out-of-box estimate for crashes involving a motor vehicle and pedestrian or cyclist

	LASSO logistic regression	Random forests	Boosting
Accuracy	0.769 [0.746 ; 0.791]	0.763 [0.739 ; 0.786]	0.753 [0.732 ; 0.774]
Sensitivity	0.785 [0.755 ; 0.815]	0.803 [0.771 ; 0.831]	0.758 [0.729 ; 0.786]
Specificity	0.749 [0.715 ; 0.783]	0.716 [0.678 ; 0.751]	0.745 [0.714 ; 0.777]
AUC	0.821 [0.801 ; 0.841]	0.812 [0.788 ; 0.836]	0.817 [0.801 ; 0.833]
Cohen kappa	0.533 [0.488 ; 0.578]	0.521 [0.474 ; 0.568]	0.502 [0.460 ; 0.544]

Responsibility prediction score

Table 7 shows the parameters of the scoring tool for crashes involving only one motor vehicle and one pedestrian or one cyclist.

Table 7: Coefficient values to construct the driver responsibility score for crashes involving 1 motor vehicle and 1 pedestrian (column 3) or 1 cyclist (column 4)

Type of variable	Variable	Value of β for crash involving 1 pedestrian	Value of β for crash involving 1 cyclist
Intercept	Intercept	$\alpha = 0.511$	
External conditions	One-way road	0.373	
	3-lane 2-way road	-1.359	1.305
	4-lane 2-way road or dual carriageway	-0.555	-1.473
	Road with separated lanes	-0.562	
	Night without road lighting	-1.011	0
	Equipped crossroads	-0.581	0
	X junction	-0.569	
	T or Y junction	0	-0.392
	Roundabout	0	2.665
	On central reservation	0	-0.894
	Off road	1.60	
	Subway, tunnel or road bridge	0	-0.749
	Steep slope	0	-0.487
	Urban road	0	-0.483
	Bad weather	0	1.254
Vehicle impact	Rear	0	-1.136
	With mobile obstacle	0	0.532
	Pedestrian masked, playing or running	0.877	*
Road-user actions	Pedestrian on pedestrian crossing	-1.071	*
	Speeding	-0.597	
	Vehicle turning left or right	0	0.289
	Insertion of vehicle in traffic	0	0.834
	Vehicle change of direction without indication	0	0.838
	Vehicle deviating left (= into traffic)	0	0.901
	Overtaking a vehicle	0	1.000
	Vehicle deviating left or right	0	1.042
	Vehicle turning left (= into traffic)**	0	1.181
	Failure to give way	0	1.462
	Changing lanes	0	2.113
	Number of faults	1.253	1.407

** not applicable for cyclists*

The score here is based on more variables than for motor vehicle crashes. Notably, more external conditions are relevant: type and form of road, and weather conditions. We can observe that for crashing involving a pedestrian, the crash happening at night decreases the responsibility score for the driver. A potential explanation is that a driver can fail to see a pedestrian because of darkness. Note that speeding is included in the number of faults, adding $1.253 - 0.597 = 0.656$ to the score for crashes involving 1 pedestrian (and $1.407 - 0.597 = 0.81$ for crashes involving 1 cyclist). While many road-user actions have no significant effect in the score assessment, the action of the pedestrian has a significant influence on the responsibility score. For the driver, variables about the pedestrian actions can be equal to 0 or -1: indeed if a pedestrian A_1 is on a pedestrian crossing, his Z variable value is $Z_1 = 1$, while the Z variable is $Z_2 = 0$ for the corresponding driver A_2. With our definition of S, the value of S for the driver becomes $S_2 = Z_2 - Z_1 = -1$. Therefore, if the pedestrian is on the pedestrian crossing, the value of -1.071 in the table becomes $S = 1.071$ for the driver. This raises the responsibility score and the driver is considered responsible unless there is a combination of at least two of the significant external factors that decrease the score. However, if the pedestrian is masked, playing or running on the road and the driver does not commit any fault, the responsibility score becomes negative and the driver will be estimated not responsible. When the crash involves a cyclist, most of the road actions of the driver significant for the responsibility score assessment are similar to the factors observed for crash configuration 1.

3.4. Single vehicle crashes (configuration 3)

There were 1,961 crashes (equal here to the number of road users) involving a single driver or riders; 1,923 drivers or riders (98.1 %) were considered responsible by the experts. We tried all three methods to predict responsibility in this configuration, but not surprisingly, the modeling proved to be very poor at predicting those not responsible. In line with the quasi-induced exposure method, our recommendation is therefore to consider all drivers involved in "single-vehicle" crashes as responsible.

4. Discussion

Estimation of responsibility in the present study comprised 3 stages.

- Driver responsibility was attributed by experts in the light of all information contained in the police reports, including crash diagrams and photographs, for a sample of about 5,000 injury crashes. Inter-observer agreement was good, suggesting that experts used similar assessment rules (Ollier & Viallon, 2014).

- Three supervised learning techniques were implemented to predict expert attribution. After cross-validation for logistic regression and boosting and out-of-bag estimation for random forests, the three methods showed similar performan*ce in terms o*f accuracy, sensitivity, specificity and reliability for crash configurations 1 and 2. We *therefore ch*ose logistic regression, which provided easy prediction based on a risk/prediction sco*re. Random* forests gave better results for single-vehicle crashes, but performance was considere*d insufficien*t for application in this case (in which drivers are considered systematically responsible*).*

- The predictors used were binary yes/no e*ncodings of in*formation in the routine police report data in France. Most of this information is al*so found in p*olice crash reports in other countries (OECD/ITF, 2016), but our objectives were obvi*ously best met* by the present data-set. However, we avoided over-fitting by using the techniques des*cribed in the* Methods section above.

The prediction score was also validated (for *the purposes* of responsibility analysis) by estimating and comparing ORs obtained for certain risk fa*ctors, using t*he predictions and expert attributions, respectively. The ORs for predictions and expert attributions were very close, except in case of high blood alcohol content, where they were lower using predictions. As drivers with high alcohol levels tend to multiply errors (Blomberg, Peck, Moskowitz, Burns, & Fiorentino, 2009), it may be that in such cases not all factors indicating extreme behaviors are fully entered in the police record database. Any such underestimation, however, would not be very important for the odds ratios, which were in any case very high, but should be taken into account in estimating the corresponding attributable risks. Brubacher et al. (2012) also found a great difference, for high blood alcohol concentrations, between their own ORs and those from large case-control studies.

The earliest driver responsibility study was conducted in Toronto by Smith (Smith & Popham, 1951), where the authors developed their own responsibility scoring tool. Their main objective was to investigate the effect of alcohol in car crashes. They used a 10-point scale to determine road-user responsibility, based on reviewing police records and distinguishing factors dependent on driver actions and those beyond their control (environmental hazards, mechanical vehicle failures). Other investigators later used responsibility analysis to study crash risks in relation to alcohol or drug use. Tehrune reviewed the limitations of using a dichotomous responsibility variable for 2-driver crashes

(Tehrune, 1983), and proposed a 5-point scale. Two inexperienced coders used his scale to determine responsibility in crash data; results suggested that there was no reason to think there would be only one responsible driver in crashes involving two or more vehicles, and that driver responsibility showed high inter-coder reliability when the responsibility was assessed on rating scales.

Later, two important studies on the same subject were published, by Robertson and Drummer (Robertson & Drummer, 1994) and, more recently, by Brubacher (Brubacher et al., 2012). A global responsibility score was constructed by attributing a-priori scores (from 1 to 5) to a series of factors presumed to increase responsibility (e.g., driver not obeying road laws, score=1) or attenuate it (e.g., vehicle hit, score=5). A global score above 15 indicates non-responsibility, 13 or less indicates responsibility, and 14 or 15 indicates undetermined responsibility.

Brubacher et al. (2012) developed an alternative prediction score based on their own expertise, implemented on a small training crash dataset containing about 100 crashes. Two experts were further asked to rate driver responsibility in the dataset: the experts' ratings and the results of the calculation were compared, and the comparison was used to improve the scoring tool.

In these two studies, to the best of our knowledge, the authors had no reference value for responsibility: validation was a-posteriori, comparing predicted scores to expert scores in a small sample.

We applied Robertson and Drummer's recommendations to our data. Comparison was, however, difficult, as we had to adapt the data to the published guidelines, with an inevitable loss of performance due to the differences between the information contained in our data set and their data, used to construct their responsibility score.

It was, however, interesting to find that applying these recommendations leads to correct predictions for crashes involving 2 or more vehicles but poorer for crashes involving cyclists or pedestrians. For crashes involving 2 or more vehicles, Cohen's kappa was 0.726 [0.708 ; 0.744], close to the value using the logistic regression model (Table 3), whereas it was only 0.159 [0.131 ; 0.187] for crashes involving a cyclist or pedestrian, as compared to 0.533 using our predictive model (Table 6).

In particular, the guidelines proposed by Robertson and Drummer and the process proposed by Brubacher take account of road surface and weather conditions for any type of crash, and conditions worse than normal reduce the risk of the driver being responsible (Brubacher et al., 2012; Robertson & Drummer, 1994). However, having a crash under bad external conditions could also be considered as a failure of the driver to adapt to the conditions: considering bad external conditions as a mitigating factor is therefore questionable (Salmi, Orriols, & Lagarde, 2014). Likewise, our score suggests that bad weather conditions increased the risk of being considered responsible only when 3 or more motor vehicles were involved.

For single vehicle crashes, we did not manage to get good performance. The very low number of non-responsible road users as assessed by the experts suggested that it is reasonable to assume that drivers were always responsible in that configuration, with some very specific exceptions where it was quite impossible for the driver to anticipate a crash situation, such as oil on the road or a truck dropping its load (Wu, Hours, & Martin, 2018).

All previous studies propose a scoring tool without explanation about the way it was assessed, in particular the values associated to each item considered. On the contrary, our present study is focused on the estimation of the driver responsibility according to a data-driven process. The parameters to be applied to each item were assessed by a machine learning algorithm, which validates the weights attributed to each predictor. Cross validation showed that our scoring tool could be applied on new crash datasets.

4.1. Limitations

The proposed score has some limitations. We decided to use a binary outcome: responsible or not responsible. There may be some crashes in which the road user's responsibility is not clearly determined. Indeed, we excluded cases in which the expert declared driver responsibility to be shared, and our score may not be efficient in such scenarios. We also did not manage to accurately identify non-responsible drivers in configuration 3 (single-vehicle crashes).

The score could directly be applicable to most of French police data, considering that drivers are almost always responsible in single vehicle crashes and that proportion of crashes with shared responsibility is very low. However it requires that all necessary variables for computing the score should be available. If the database does not have such information, the score may not give a reliable determination of responsibility. However, the methodology could also be adapted for other national police data, and R scripts are available from the authors upon request.

Further work is needed to validate this responsibility assessment, notably using similar police data such as those in the European CARE database, which is the Community database on road crashes resulting in death or injury, comprising detailed data on individual crashes as collected by the Member States.

5. Conclusion

Responsibility analysis enables crash risk factors to be quantified, given certain hypotheses (Brubacher et al., 2014), without resort to exposure data, which is why it is widely used (Salmi et al., 2014). Results greatly depend on the quality of how responsibility is determined, and it is equally important that the elements used for determination should be explicit, allowing interpretation of identified risk factors.

Based on expert decisions for a fairly large number of police crash reports, we constructed a score to assess responsibility for drivers and riders in crashes involving one or more motor vehicles, or involving a cyclist or pedestrian. Odds ratios estimated from the score were similar to those estimated from expert assessment, and cross-validation showed that it can also predict expert responsibility assessments on new data sets.

We believe that this score can be used to reliably assess responsibility based on national police report databases, provided that they include the information needed to construct the score. It can then be used to perform responsibility analysis to identify and study transient and stable risk factors for road crashes.

References

af Wåhlberg, A. E., & Dorn, L. (2007). Culpable versus non-culpable traffic accidents; What is wrong with this picture? *Journal of Safety Research*, *38*(4), 453–459. https://doi.org/10.1016/j.jsr.2007.01.013

Blaizot, S., Papon, F., Haddak, M. M., & Amoros, E. (2013). Injury incidence rates of cyclists compared to pedestrians, car occupants and powered two-wheeler riders, using a medical registry and mobility data, Rhone County, France. *Accid Anal Prev*, *58*, 35–45. https://doi.org/10.1016/j.aap.2013.04.018

Blomberg, R. D., Peck, R. C., Moskowitz, H., Burns, M., & Fiorentino, D. (2009). The long beach/fort lauderdale relative risk study. *Journal of Safety Research*, *40*(4), 285–292.

Bouaoun, L., Haddak, M., & Amoros, A. (2015). Fatal road traffic crashes: comparisons by road user types and measures of exposure. *Accid Anal Prev*, *75*, 217–225.

Brubacher, J., Chan, H., & Asbridge, M. (2012). Development and validation of a crash culpability scoring tool. *Traffic Injury Prevention*, *13*(3), 219–229. https://doi.org/10.1080/15389588.2011.645383

Brubacher, J., Chan, H., & Asbridge, M. (2014). Culpability analysis is still a valuable technique. *International Journal of Epidemiology*, *43*(1), 270–272. https://doi.org/10.1093/ije/dyt142

Chandraratna, S., & Stamatiadis, N. (2009). Quasi-induced exposure method: Evaluation of not-at-fault assumption. *Accident Analysis & Prevention*, *41*(2), 308–313. http://dx.doi.org/10.1016/j.aap.2008.12.005
Cooper, P. J., Meckle, W., & Andersen, L. (2010). The efficiency of using non-culpable crash-claim involvements from insurance data as a means of estimating travel exposure for road user sub-groups. *Journal of Safety Research*, *41*(2), 129–136. https://doi.org/10.1016/j.jsr.2010.02.004
Davis, G. A., & Gao, Y. (1995). Statistical methods to support induced exposure analyses of traffic accident data. *Transportation Research Record*, *1401*, 43–49.
Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Ann Stat*, *32*(2), 407–499.
Haight, F. A. (1973). Induced exposure. *Accident Analysis & Prevention*, *5*(2), 111–126. http://dx.doi.org/10.1016/0001-4575(73)90019-5
Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The elements of statistical learning* (2nd ed.). Springer.
Lyles, R. W., Stamatiadis, P., & Lighthizer, D. R. (1991). Quasi-induced exposure revisited. *Accident Analysis & Prevention*, *23*(4), 275–285. http://dx.doi.org/10.1016/0001-4575(91)90005-P
OECD/ITF. (2016). *Road Safety Annual Report 2016* (OECD Publishing). Paris. Retrieved from http://dx.doi.org/10.1787/irtad-2016-en
Ollier, E., & Viallon, V. (2014). Joint estimation of K related regression models with simple L_1 -norm penalties. *ArXiv Preprint ArXiv:1411.1594*.
Robertson, M. D., & Drummer, O. H. (1994). Responsibility analysis: A methodology to study the effects of drugs in driving. *Accident Analysis & Prevention*, *26*(2), 243–247. https://doi.org/10.1016/0001-4575(94)90094-9
Sagberg, F. (2001). Accident risk of car drivers during mobile telephone use. *International Journal of Vehicle Design*, *26*, 57–69.
Salmi, L. R., Orriols, L., & Lagarde, E. (2014). Comparing responsible and non-responsible drivers to assess determinants of road traffic collisions: time to standardise and revisit. *Injury Prevention*, *20*(6), 380–386. https://doi.org/10.1136/injuryprev-2013-041143
Smith, H. W., & Popham, R. E. (1951). Blood alcohol levels in relation to driving. *Can Med Assoc J*, *65*(4), 325–328.
Stamatiadis, N., & Deacon, J. A. (1997). Quasi-induced exposure: Methodology and insight. *Accident Analysis & Prevention*, *29*(1), 37–52. http://dx.doi.org/10.1016/S0001-4575(96)00060-7
Tehrune, K. W. (1983). An evaluation of responsibility analysis for assessing alcohol and drug crash effects. *Acc Anal Prev*, *15*, 237–246.
Thorpe, J. (1964). Calculating relative involvment rates in accidents without determining exposure, *2*, 25–36.
Wu, D., Hours, M., & Martin, J.-L. (2018). Risk factors for motorcycle loss-of-control crashes. *Traffic Injury Prevention*, *19*(4), 433–439. https://doi.org/10.1080/15389588.2017.1410145

Appendix A. Scoring tool examples

The first example illustrates how to use the score for a crash between two motor vehicles, as derived from configuration 1. Here we calculate the score of driver A. In this example, driver B was above the speed limit, and failed to give way to driver A on the right at an intersection. Driver B's vehicle received impact on the right, and driver A's vehicle received frontal impact.

	Variable	W(A)	W(B)	
External conditions	One-way road	0	0	
	4-lane 2-way road or dual carriageway	0	0	
	Main road	0	0	
	Urban road	0	0	
	Bridge, tunnel or subway	0	0	
	Heavy rain	0	0	
	Fog, snow or storm	0	0	
	Variable	Z(A)	Z(B)	S(A)
Vehicle impact	Rear	0	0	0
	Frontal	1	0	1
	Right	0	1	-1
	With mobile obstacle	1	1	0
	With fixed obstacle	0	0	0
Driver actions	Stationary vehicle	0	0	0
	Speeding	0	1	-1
	Avoidance maneuver	0	0	0
	Turning left*	0	0	0
	Overtaking on left*	0	0	0
	Vehicle deviating left*	0	0	0
	Vehicle turning left or right	0	0	0
	Insertion of vehicle in traffic	0	0	0
	Overtaking	0	0	0
	Between lanes, half-turn or reversing	0	0	0
	Changing lanes	0	0	0
	Vehicle deviating	0	0	0
	Number of faults	0	2	-2

*assuming a country in which driving is on the right and a lefthand-drive vehicle.

The score for driver A is:

$$R(A) = 0.023 + 1 \times 0 + (-1) \times 0.062 + (-1) \times 0 + (-2) \times 1.897 = -3.83$$

Since $R(A) < 0$, driver A is predicted as non-responsible for the crash.

Accepted for publication in Journal of Safety Research on July 2, 2019 https://doi.org/10.1016/j.jsr.2019.07.001

A second example is given to illustrate how to use the score for a crash between a car A and a pedestrian B. Urban area, at an X junction, car turning left, hits a pedestrian on a pedestrian crossing.

	Variable	W(A)	W(B)	
External conditions	One-way road	0	0	
	3-lane 2-way road	0	0	
	4-lane 2-way road or dual carriageway	0	0	
	Separated lanes	0	0	
	Night without lighting	0	0	
	Equipped crossroads	0	0	
	X junction	1	1	
	T or Y junction	0	0	
	Roundabout	0	0	
	On central reservation	0	0	
	Off road	0	0	
	Bridge, tunnel or subway	0	0	
	Steep slope	0	0	
	Urban road	1	1	
	Bad weather	0	0	
	Variable	Z(A)	Z(B)	S(A)
Vehicle impact	Rear	0	0	0
	With mobile obstacle	0	0	0
Road-user actions	Pedestrian masked, playing or running	0	0	0
	Pedestrian on pedestrian crossing	0	1	-1
	Speeding	0	0	0
	Vehicle turning left or right	1	0	1
	Insertion of vehicle in traffic	0	0	0
	Change of direction without indicating	0	0	0
	Vehicle deviating left*	0	0	0
	Overtaking	0	0	0
	Vehicle deviating left or right*	0	0	0
	Vehicle turning left*	1	0	1
	Failure to give way	0	0	0
	Changing lanes	0	0	0
	Number of faults	0	0	0

*assuming a country in which driving is on the right and a lefthand-drive vehicle.

The score for driver A is:

$$R(A) = 0{,}511 + 1 \times (-0{,}569) + 1 \times 0 + (-1) \times (-1{,}071) + 1 \times 0 + 1 \times 0 = 1{,}01 > 0$$

Driver A is therefore predicted to be responsible.