KINEMATICAL, PHYSIOLOGICAL, AND VEHICLE-RELATED INFLUENCES ON PEDESTRIAN INJURY SEVERITY IN FRONTAL VEHICLE CRASHES: MULTIVARIATE ANALYSIS AND CROSS-VALIDATION

Dipl.-Ing. Thomas Helmer[(1)], Randa Radwan Samaha[(2)], M.Sc., Paul Scullion[(2)], B.Sc., Dipl.-Ing. Adrian Ebner[(1)], Dr. Ronald Kates[(3)]
[(1)]BMW AG, [(2)]George Washington University, [(3)]REK Consulting

ABSTRACT

Evaluation of safety benefits is an essential part during design and development of pedestrian protection systems. Translating physics into human benefits for large-scale simulations of the benefit requires reliable and validated injury and fatality models. To this end, multivariate predictive models for pedestrian fatalities and different injury severities by means of the ISS as well as the MAIS scale are developed using the US Pedestrian Crash Data Study (PCDS). In addition to impact speed, significant multivariate predictors include physiological and vehicle characteristics. The in-sample as well as out-of-sample predictive quality is remarkably high. An approach to define a metric suitable for comparing active and passive safety is presented. As active safety is capable of influencing impact speed directly, the benefits of a reduction of impact speed regarding injury probability / mortality while controlling other influencing factors is computed. The relative reduction of injury probability / mortality with respect to exposure seems to be an appropriate metric to evaluate and compare safety benefits of measures considering both active and passive safety.

KEYWORDS: ACCIDENT ANALYSIS, PEDESTRIAN, INJURY PROBABILITY, AUTOMOBILES, STATISTICS

One of the great challenges in motor vehicle safety is the protection of vulnerable road users. The protection of pedestrians as being the most vulnerable participants in traffic is of special interest as it is gaining more and more attention in the scientific community as well as in legislation and consumer protection [Nations, 2009, Union, 2009, Wisselmann, 2009, NCAP, 2010]. Different measures of passive safety approaches, such as changes in vehicle design or active elements, such as the so-called “pop-up hood”, are being introduced in the market by car manufacturers [Friesen, 2003, Inomata, 2009].

Approaches of active safety for pedestrian protection are currently being developed in the automobile industry and close to market introduction. Active or preventive pedestrian protection systems in the vehicle usually detect pedestrians and interact with the driver or automatically intervene in case of an imminent collision. This results in a reduction in impact speed which mitigates the impact or even avoids the collision entirely.

The efficacy of such technologies has to be evaluated and the results have to be presented in an understandable way using a common metric for objective evaluation as well as for comparison between measures of active and passive safety. Simulation techniques which include both deterministic and stochastic processes are currently being developed and utilized for the optimization of pedestrian protection systems. Those “Monte-Carlo” simulation methods typically simulate millions of pedestrian situation, a small fraction of which lead to unfavorable outcomes such as collisions. By virtually implementing a safety measure, the shift in frequency distribution as well as impact characteristics show the effect of that safety measure.

To evaluate the overall efficacy relative reduction of injury severity of the pedestrian is clearly an appropriate metric, because it addresses direct human costs. The HARM methodology [Gabler, 2005] can be applied in addition in order to highlight the overall social and economic impacts of a crash in particular or of measures of pedestrian protection in general. It includes direct medical treatment and rehabilitation costs, but also considers reduction in productivity based on injury severity. An interface between the simulation-based characteristics of the impact and resulting injuries as well as costs needs to be developed in order to evaluate the overall benefit of a given safety measure.

This paper presents statistical methodology in order to derive models for predicting injury severity / mortality of the pedestrian involved in frontal vehicle crashes, based on the conditions of impact. The models provide a conditional probability of each injury level for any given set of impact characteristics. Collision speed is by far the most important factor [Hannawald, 2004, Rosen, 2009b]; however, additional factors of the pedestrian, such as age or physiological variables, as well as vehicle related factors are also relevant in the multivariate models.

This paper also considers the question of what dependent variables, i.e. the Maximum of the Abbreviated Injury Scale (MAIS) [States, 1980], the Injury Severity Score (ISS) [Baker, 1974] or the New Injury Severity Score (NISS) [Osler, 1997], are most suitable for characterizing injuries. The results are intended to improve and quantify the predictability of pedestrian injuries in accidents and to provide an interface between simulation and detailed evaluation of active pedestrian safety systems in terms of human benefits.

DATA AND METHODS

STUDY DATA CHARACTERISTICS

The US Pedestrian Crash Data Study (PCDS) was used for this study and is characterized as follows [Helmer, 2010]. PCDS contains a sample of pedestrian accidents between 1994 and 1998, generated by the Transportation Data Center at the University of Michigan's Transportation Research Institute for the National Center for Statistics and Analysis (NCSA) of the National Highway Traffic Safety Administration (NHTSA). A total of 552 cases in 6 US cities were collected. Most of the cases were not included in the crashworthiness data system (CDS) or the general estimates system (GES). The main selection and exclusion criteria for PCDS may be summarized as follows:

- Pedestrian accidents only (no cyclists, etc).
- Impact between one vehicle and one pedestrian.
- Cars and light trucks only.
- Vehicles of model year 1990 or later.
- Initial contact of the vehicle with the pedestrian was in front of the A-pillar.
- Vehicle part striking pedestrian was (undamaged) original equipment.

Due to the sampling scheme and selection criteria, the PCDS data set may not be entirely representative for all pedestrian accidents in the US, but it is quite useful for the intended purpose of identifying risk factors and estimating predictive risk models in the accident classes considered [UMTRI, 2005, Isenberg, 1998, Jarrett, 1998]. However, PCDS seems to be quite representative considering the frequency distribution of accident scenarios [Ebner, 2010].

Data were filtered for frontal vehicle impact with the pedestrian for ages four and older, resulting in 450 collisions. For the following analysis, only cases with available impact speed were considered, as this is by far the most important factor for injury severity of the pedestrian, resulting in 376 collisions.

CODING OF VARIABLES AND TREATMENT OF MISSING DATA

Selection and coding of target variables

Pedestrian injuries were originally reported according to the maximum of the Abbreviated Injury Scale (MAIS), revision 90 [States, 1969, States, 1980, AAAM, 1990] and the Injury Severity Score (ISS), both of which are ordinal measures of severity [Baker, 1976, Baker, 1974, States, 1980, States, 1969]. Meaningful levels for MAIS as binary outcome variables are MAIS2+ (MAIS≥2, at least moderate injury), MAIS3+ (MAIS≥3, at least serious injury), MAIS5+ (MAIS≥5, at least critical injury) [Hannawald, 2004, Kühn, 2005]. MAIS4+ and MAIS6 were coded but not further analyzed.

Following [Hakkert, 2007] in the analysis, this study focuses on ISS and binary outcome variables are coded as ISS9+ (ISS≥9, at least serious injury) ISS16+ (ISS≥16, at least serious injury), ISS25+ (ISS≥25, at least critical injury).

The ISS is calculated as the sum of the squares of the highest AIS scores in each of the three most severely injured body regions (out of 6 regions). It ranges from 0 to 75; 75 is defined if at least one body region has an AIS of 6. There are indications that ISS gives a more precise estimate of the overall injury severity than MAIS [Nogueira, 2008, Stevenson, 2001]. The applicability as well as the changes and challenges of AIS, ISS (e.g., the possibility of obtain a certain ISS by different combination of AIS values) and NISS are extensively discussed in medical literature. The consequences for this study will be discussed while comparing multivariate models and their in-sample predictive performance for MAIS and ISS as outcome categories.

The consistency of the data regarding MAIS and ISS was checked by computing the ratio of squared MAIS by ISS which must be within 0.33 (i.e., three body regions have an injury which severity equals the MAIS of that person) and 1.0 (i.e., only one injury per person).

The new ISS (NISS), which is calculated identically to the ISS but uses the three single injuries with the highest AIS, is supposed to be an improvement to ISS [Nogueira, 2008, Osler, 1997]. As it is not included in PCDS, it was additionally calculated for the database and compared to the ISS. As 95% of the values are identical in the whole database, an investigation of NISS doesn't seem to be reasonable.

Fatalities were coded and investigated independently, as they are distributed over a range of ISS or MAIS values [Rosen, 2009a]. Missing injury data are not imputed as they are used as target variables.

Splitting the population into subgroups by age is sound for biomechanical reasons and is investigated in [Helmer, 2010]. To achieve manageable statistical power considering limited case numbers this study focuses on ages 18+ (18 or greater) but still treats age as an independent predictive factor within this group. Table 1 summarizes the frequencies for each outcome category. Numbers in parenthesis refer to the number of cases not in the outcome category.

Table 1. Frequency for Each Outcome Category

Injury level	MAIS 2+	MAIS 3+	MAIS 5+	ISS 9+	ISS 16+	ISS 25+	Fatalities
Age 18+	165 (120)	119 (166)	46 (239)	128 (158)	102 (184)	78 (208)	45 (241)

Coding of explanatory variables and treatment of missing data

The coding and imputation of variables is identical to the one used in [Helmer, 2010]. Explanatory variables considered in this paper fall into the following general categories:

- Vehicle kinematics, including impact speed.
- Vehicle profile characteristics (static).
- Driver maneuvers and attention.
- Pedestrian physiology, including ratios.
- Pedestrian movement.

Imputation of missing values was carried out (using secondary data resources as far as available) in order to ensure sufficient case numbers within the 376 cases (see above). Body height and weight were imputed using anthropometric data from the National Health and Nutrition Examination Survey (NHANES) [Ogden, 2004]. As mentioned above impact speed has not been imputed due to the effect as most important factor for injury causation. The possible effects due to the accuracy of accident reconstruction are not discussed here in detail, but have been considered.

Imputation does narrow the resulting distributions of explanatory variables and leads to an underestimate of the variance in logistic regression. Nonetheless, it is far preferable to the usual procedure of list-wise deletion both by avoiding a loss of statistical power and by minimizing biases.

Continuous variables (with the exception of impact speed) were transformed by subtracting the mean and dividing by the standard deviation (SD) computed from the full sample in order to have a comparable and uniform basis for further analysis (see Appendix Table 12 for a complete list of continuous variables and transformations). There was no significant difference, considering the mean between the full and the subset with impact speed with respect to age, body height, body mass index (BMI) or ground-to-forward hood opening. Several ratios of vehicle dimensions to anthropometric values, squared variables (height, impact speed) as well as combined values (kinetic energy, BMI) were also defined and transformed. Impact speed was scaled by the mean. The square of this ratio was also coded as a separate explanatory variable to account for nonlinear effects (in addition to kinetic energy).

Categorical variables were recoded as binaries by considering each category separately. For example, for “attempted avoidance maneuvers of the car”, the combinations of steering, braking, and accelerating were coded separately as new binary variables. Ordinal (non-continuous) variables were recoded as cumulative binary quantities. A binary variable distinguishing light trucks (1) from passenger cars (0) was also defined, where the group of automobiles is separated from the rest (utility vehicles and trucks). (See Appendix Table 15 and Table 16 for a complete list of non-continuous variables and transformations).

STATISTICAL METHODS

Spearman correlations, t-tests, and Mann-Whitney tests were performed to assess internal relationships among explanatory variables. PASW Statistics and Microsoft Office Excel were used for computations.

Univariate and multivariate binary logistic regression was used to determine unadjusted and adjusted odds ratios (respectively) and to construct risk scores for the injury severities and for fatalities as explained above. Accordingly, the probability p_i for a pedestrian to reach or exceed a certain level of overall injury severity can in principle be obtained from a model via the formula

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{1,i} + \ldots + \beta_k x_{k,i})}{1 + \exp(\beta_0 + \beta_1 x_{1,i} + \ldots + \beta_k x_{k,i})} \quad , \qquad (1.)$$

where $x_{1,i} \ldots x_{k,i}$ are explanatory factors for the collision such as speed, vehicle dimensions, etc., and the coefficients $\beta_1 \ldots \beta_k$ are those estimated in the regression process.

The probability estimate obtained from a model of the form Eq. (1) may also be thought of as a “risk score”. The quantity [$\exp(\beta_j)$] resulting from multivariate regression is known as the “adjusted odds ratio” of the explanatory factor with the label *j* for the outcome in question. If only one factor, say x_1, is entered into binary regression, the regression is referred to as “univariate” and the quantity [$\exp(\beta_1)$] is then known as the “crude” or “unadjusted” odds ratio for the factor x_1.

Variables with suspected impact on injury severity were first tested for univariate impact; multivariate logistic regression models were constructed for the subgroups mentioned above and evaluated for the binary injury endpoints of interest. Model selection in our multivariate models was performed by standard forward elimination, using the likelihood ratio statistic. Factors that fail to be significant in a particular multivariate model are regarded as associated with a β coefficient of zero or equivalently with an odds ratio of one. Failure to reach significance in this context does not necessarily mean that a factor is truly irrelevant, but simply that it is not possible to reject the null hypothesis at the assumed level of significance. The 95% confidence intervals of odds ratios give an indication of the validity of the findings. In the case of an odds ratio, a significant p-value ($p<0.05$) is equivalent to the statement that the 95% confidence region does not include the value one.

In order to assess and compare the predictive quality of models containing different numbers of explanatory factors, we evaluated the well-known Akaike and Bayes information criteria

$$AIC = -2 * LL + 2 * (k+1) \tag{2.}$$

$$BIC = -2 * LL + \ln(n) * (k+1) \tag{3.}$$

where LL is the log-likelihood, k the number of model parameters, and n the number of cases. Lower values indicate improved model fit. The absolute values of AIC and BIC have no meaning; only relative differences help us rank different models and indicate which has the best fit among the given alternatives. The BIC is very closely related to AIC, but it has a penalty term for the number of parameters in a model.

Using risk scores according to Eq. (1), the estimated benefit of an improvement in a significant causal factor (e.g., impact speed) on injury probability (ISS9+, etc.) may be quantified by comparing the injury probability with versus without the improvement. It is also instructive to compare the effects of varying different causal factors on relative risk changes. Note that the comparison is made holding all other significant factors constant.

Issues of confounding and multicollinearity, which nearly always occur in observational data sets, will be addressed as they arise. Multicollinearity in the present context refers to the fact that within a multivariate model, β regression coefficients of highly correlated explanatory factors are interdependent. Thus, the apparent predictive impact of one factor can depend on whether or not a distinct but correlated factor is included in the analysis or attains significance.

For each model, the areas (AUC) under the receiver operating characteristics (ROC) were also evaluated as indicators for in-sample predictive accuracy. The model quality indicates the degree to which injury severity is predictable from measurable factors included in the in-depth accident data set. Theoretically, if the ROC-AUC of a statistical model is 1.0, it is a perfect predictor; if the ROC-AUC is not significantly different from 0.5, then the factors have no predictive value at all (e.g., if all factors are eliminated in logistic regression).

10-fold cross validation was used to evaluate the predictive accuracy as well as the statistical question of robustness of the models. The underlying question addresses the general validity and reliability of statements based on models derived from limited data sets. In 10-fold cross validation, the logistic regression model is repeatedly evaluated using nine-tenths of the data for training and one-tenth for assessment. This procedure represents an important first step toward estimating a realistic out-of-sample predictive power. The stability of the variables in the full-data models is also assessed by their frequency of occurrence in cross-validation models.

ROC-AUC was found to be an appropriate statistic to quantify the amount of optimism in the models. We also considered false classification rate with a risk-adjusted cutoff as a statistic for the cross-validation. The cutoff value is the number of cases not being in the outcome category in relation to the total number of cases. If the predicted risk is greater than the cutoff value, the case is classified as 1 (else 0). The false classification rate seems to be less suited as a statistic for this study as it is highly sensitive to the low number of cases in each test group and therefore produces high variance.

RESULTS

UNIVARIATE AND MULTIVARIATE ANALYSIS

Univariate results

Table 2 and Table 3 summarize the univariate impact of all explanatory variables that were significant predictors (p-value ≤ 0.05), for serious or worse injuries (ISS9+) as well as for fatalities (due to space available, no other univariate results are displayed). As expected, impact speed is the strongest single predictive factor (either as a scaled variable, squared or as part of kinetic energy). In order to interpret the odds ratios of these tables, note that, taking into account the scaling of continuous quantities defined here, the additional risk associated with a significant odds ratio >1 (or decrease associated with an odds ratio <1) refers to a one-standard deviation change in the factor. For example, in Tables 2-3, odds ratios >1 are associated with impact speed: 16.3 (ISS9+); 8.5 (fatalities) respectively. These odds ratios each refer to a hypothetical increase in collision speed equal to the scaling factor (28.9 km/h). The odds ratios associated with an one-half as large (14.5 km/h) increase in speed would be estimated as the *square root* of these respective odds ratios in each case. (This statement follows from Eq. (1) and the definition of an odds ratio.)

Table 2 . Univariate Results for ISS9+ (age group 18+)

Variable	N	n ISS 0-8	n ISS 9+	Scaling factor	p-value	unadjusted odds ratio	Confidence interval		AIC	BIC
Vehicle kinematics (incl. impact speed)										
Impact speed	286	158	128	28.9	< 0.001	16.325	8.400	31.727	264	271
Kinetic energy	286	158	128	113.1	< 0.001	38.786	13.995	107.496	264	271
Impact speed (squared)	286	158	128	$(28.9)^2$	< 0.001	3.849	2.652	5.584	269	277
Vehicle characteristics (static)										
Hood length	286	158	128	19.5	0.006	1.413	1.104	1.810	389	397
Front bumper lead	285	157	128	3.0	0.004	1.445	1.128	1.850	387	394
Ground to rear hood opening (wrap)	286	158	128	22.2	0.003	1.432	1.129	1.818	388	396
Ground to base of windshield (wrap)	286	158	128	22.6	0.003	1.439	1.129	1.832	388	395
Ground to top of windshield (wrap)	286	158	128	21.1	0.004	1.449	1.125	1.865	388	396
Driver maneuvers and attention										
Avoidance: Steering right	286	158	128	-	0.010	7.333	1.595	33.717	388	395
Avoidance: Steering	286	158	128	-	0.002	3.123	1.543	6.323	387	394
Pre-event movement car: complexity	286	158	128	-	< 0.001	0.193	0.112	0.333	358	365
Pedestrian physiology (incl. ratios)										
Age (pedestrian)	286	158	128	22.2	0.010	1.451	1.093	1.928	391	398
Body mass index	286	158	128	5.7	0.035	1.322	1.020	1.715	393	400
Hip height (ped., imputed)	286	158	128	11.1	0.040	0.651	0.432	0.981	393	400
Ratio height shoulder / ground rear hood op.	286	158	128	0.1	0.021	0.710	0.531	0.950	392	399
Ratio body height / ground rear hood opening	286	158	128	0.2	0.032	0.729	0.546	0.974	392	400
Ratio body height / ground base windshield	286	158	128	0.1	0.043	0.734	0.544	0.991	393	400
Ratio body height / hood length	286	158	128	0.5	0.036	0.759	0.587	0.982	392	400
Sex	286	158	128	-	0.038	0.607	0.379	0.972	393	400
Pedestrian movement										
Walking: speed	286	158	128	-	0.043	1.642	1.016	2.652	393	401
Walking: danger	286	158	128	-	0.032	1.678	1.045	2.693	393	400

Several geometric characteristics are associated with higher risk in both outcome categories, all with the expected trend. Among the physiology of the pedestrian age and BMI were significant. Of interest is the finding that an increasing ratio from body height to vehicle geometry is associated with decreasing risk for ISS9+ injuries as well as fatalities.

Table 3. Univariate Results for Fatalities (age group 18+)

Variable	N	n not fatal	n fatal	Scaling factor	p-value	unadjusted odds ratio	Confidence interval		AIC	BIC
Vehicle kinematics (incl. impact speed)										
Impact speed	286	241	45	28.9	< 0.001	8.501	4.749	15.217	167	175
Kinetic energy	286	241	45	113.1	< 0.001	4.177	2.642	6.604	184	191
Impact speed (squared)	286	241	45	$(28.9)^2$	< 0.001	1.923	1.588	2.330	174	182
Vehicle characteristics (static)										
Hood length	286	241	45	19.5	0.024	1.542	1.058	2.245	247	254
Front bumper lead	285	240	45	3.0	0.016	1.499	1.078	2.085	246	254
Front-top transition point (height)	286	241	45	15.6	0.017	1.414	1.063	1.881	247	255
Ground to rear hood opening (wrap)	286	241	45	22.2	0.002	1.650	1.193	2.283	243	251
Ground to base of windshield (wrap)	286	241	45	22.6	0.002	1.653	1.208	2.261	243	250
Ground to top of windshield (wrap)	286	241	45	21.1	0.013	1.520	1.094	2.111	247	254
Driver maneuvers and attention										
Avoidance: Steering left	286	241	45	-	0.055	2.389	0.980	5.822	250	257
Pre-event movement car: complexity	286	241	45	-	< 0.001	0.203	0.083	0.497	237	244
Pedestrian physiology (incl. ratios)										
Age (pedestrian)	286	241	45	22.2	0.003	1.772	1.223	2.567	244	251
Weight (ped., imputed)	286	241	45	22.6	0.001	1.890	1.283	2.784	243	250
Body mass index	286	241	45	5.7	< 0.001	1.745	1.279	2.381	240	248
Ratio height hip / front-top transition	286	241	45	0.2	0.009	0.621	0.435	0.887	246	253
Ratio height shoulder / ground rear hood op.	286	241	45	0.1	0.006	0.541	0.349	0.839	245	252
Ratio body height / front-top transition	286	241	45	0.4	0.024	0.665	0.467	0.947	248	255
Ratio body height / ground rear hood opening	286	241	45	0.2	0.010	0.561	0.362	0.870	246	253
Ratio body height / ground base windshield	286	241	45	0.1	0.009	0.553	0.355	0.860	245	253
Pedestrian movement										
Direction ped.: with / against traffic	286	241	45	-	0.012	4.896	1.426	16.804	247	255
Direction ped.: towards lane / crossing	286	241	45	-	0.009	0.285	0.112	0.727	247	254

Analysis of potential confounders and associations between explanatory variables

Several variables associated with driver or pedestrian movement were significant in the univariate analyses above. These include avoidance maneuvers by the driver (steering) as well as the walking direction of the pedestrian, his walking speed and the danger of his maneuver (standing or running is associated with higher danger whereas walking is considered less critical). Prior vehicle movements are grouped by their complexity for the driver (e.g. turning, merging, lane changing, etc. are regarded as complex). The causal relationship of these variables to injury severity could be complex.

As impact speed is such a dominant determinant for injury or fatality, any explanatory variable in the database that is associated with impact speed could act as a surrogate for impact speed and thus as a confounder, i.e., being significant without having a causal relationship. These potential confounders were tested for association using t- and Mann-Whitney-Tests. P-Values below refer to the hypothesis of differences in impact speed for the two groups defined by the binary variable discussed.

For example the unadjusted odds ratio associated with walking speed of the pedestrian (see Table 2 and Table 3) can be explained with significantly lower impact speed of the vehicle among those moving slowly (26.3 vs. 32.0 km/h, $p<0.001$). Consistently higher walking danger corresponds with significantly ($p<0.001$) higher impact speed (32.3 vs. 25.6 km/h). Evasive steering of the driver is associated with higher collision speeds (38.9 vs. 27.3 km/h, $p<0.001$). It can be assumed that evasive steering is more likely in critical situations with higher speeds, where the driver senses that the time to collision is not adequate for effective mitigation by braking. Complexity of the pre-event maneuver of the vehicle (17.6 vs. 35.0 km/h, $p<0.001$) as well as walking direction of the pedestrian (28.0 vs. 39.9 km/h, $p<0.001$, for crossing pedestrians vs. not crossing) are also surrogates for impact speed [Helmer, 2010].

Multivariate results

Based on the univariate results, variables describing vehicle kinematics, vehicle characteristic, and pedestrian physiological characteristics were entered into multivariate models, but not variables describing pre-event maneuvers or pedestrian movement, even if significant in univariate analysis.

Variables describing pre-event maneuvers or pedestrian movement can indeed show a statistical association with injury severity, for example due to selection effects (i.e., they determine what crashes actually occurred). However, for the purposes of the present study, one needs to consider factors that influence pedestrian injuries (i.e., for biomechanical or physical reasons), conditional on the crash actually occurring. For this reason, variables describing pre-event maneuvers or pedestrian movement were omitted from the multivariate analysis.

The results were computed for binary outcome variables (ISS9+, etc.). The multivariate models are segregated as explained above with respect to age groups and presented in Tables 4-6.

Table 4. Multivariate Results for Fatalities (age group 18+)

Variable	Scaling factor	p-value	adjusted odds ratio	Confidence interval		AIC	BIC
Age group 18+						137	159
Body mass index	5.7	0.001	2.147	1.393	3.308		
Impact speed	28.9	< 0.001	62.001	11.647	330.051		
Kinetic energy	113.1	0.019	0.360	0.154	0.843		
Age (pedestrian)	22.2	0.004	2.220	1.288	3.828		
Ratio height hip / front-top transition	0.2	0.008	0.405	0.207	0.794		
$\exp(\beta_0)$			0.000				
Age group 18+ (only speed)						167	175
Impact speed	28.9	0.297	8.501	4.749	15.217		
$\exp(\beta_0)$			0.011				
Age group 18+ (only speed, age)						156	167
Impact speed	28.9	0.314	9.688	5.239	17.913		
Age (pedestrian)	22.2	0.256	2.458	1.488	4.060		
$\exp(\beta_0)$			0.005				

The resulting models contain up to 5 independent variables. As in univariate analysis, impact speed is the strongest single predictive factor in every model. Age of the pedestrian is also included in every model and is positively associated with increased risk. In the model regarding fatalities kinetic energy is entered as significant factor. Higher kinetic energy is associated with decreasing risk whereas univariate kinetic energy has a high unadjusted odds ratio. This effect can be explained by a very strong presence of linear impact speed in the model. Since kinetic energy contains squared impact speed this adjusted odds ratio could level off linear fatality risk at high impact speed or considers other non-linearities in the dependence on impact speed.

Note that higher ground to hood opening (which describes the vertical distance from the ground to the forward hood opening and which is comparable to the transition point between front and top of the hood) also corresponds to increased risk. Higher ground to hood opening and longer ground to base of windshield are characteristics of light trucks and sports utility vehicles. Due to the grouping mentioned above, passenger cars and light trucks show no univariate significance in this study. However, light trucks, utility vehicles, and van based trucks result in a lower probability for MAIS3+ of the pedestrian than conventional trucks (for the higher injury risk associated with light truck see also [Dahdah, 2008, Henary, 2003, Scullion, 2010, Lefler, 2001, Kerrigan, 2008, Lefler, 2004, Roudsari, 2005, Simms, 2006, Longhitano, 2005]).

Table 5. Multivariate Results for Injuries Using ISS (age group 18+)

Variable	Scaling factor	p-value	adjusted odds ratio	Confidence interval		AIC	BIC
ISS 9+ (age group 18+)						254	265
Impact speed	28.9	< 0.001	20.058	9.595	41.930		
Ground to base of windshield (wrap)	22.6	0.026	1.413	1.043	1.914		
Age (pedestrian)	22.2	0.001	1.888	1.285	2.773		
exp(β_0)			0.039				
ISS 9+ (age group 18+) speed						264	271
Impact speed	28.9	< 0.001	16.325	8.400	31.727		
exp(β_0)			0.057				
ISS 16+ (age group 18+)						211	225
Impact speed	28.9	< 0.001	31.045	13.630	70.712		
Forward hood opening (height)	17.0	0.005	1.539	1.139	2.080		
Age (pedestrian)	22.2	0.002	1.956	1.280	2.990		
exp(β_0)			0.012				
ISS 16+ (age group 18+) speed						225	232
Impact speed	28.9	< 0.001	21.017	10.349	42.682		
exp(β_0)			0.023				
ISS 25+ (age group 18+)						159	181
Weight vehicle	34.1	0.025	1.787	1.076	2.970		
Impact speed	28.9	< 0.001	47.117	17.606	126.095		
Forward hood opening (height)	17.0	0.038	1.736	1.030	2.925		
Age (pedestrian)	22.2	< 0.001	2.559	1.528	4.284		
Bumper leading angle	15.4	0.002	0.474	0.293	0.767		
exp(β_0)			0.002				
ISS 25+ (age group 18+) speed						189	196
Impact speed	28.9	< 0.001	21.822	10.493	45.381		
exp(β_0)			0.010				

Table 4 and Table 5 give also univariate models with only impact speed as predictive variable as they represent the current state of the art [Hannawald, 2004]. These models will help in evaluating the improvement in predictive accuracy due to adding additional factors (see sections below).

Another question is the comparison of ISS and MAIS as outcome category for overall injury severity while used in logistic regression models based on in-depth accident (i.e., observational) data. To this end, Table 6 gives the multivariate results for MAIS2+, MAIS3+, and MAIS 5+. Each model again contains impact speed as dominant predictive factor. One additional factor is entered in each model.

Table 6. Multivariate Results for Injuries Using MAIS (age group 18+)

Variable	Scaling factor	p-value	adjusted odds ratio	Confidence interval		AIC	BIC
MAIS 2+ (18+)						285	296
Impact speed	28.9	< 0.001	11.722	5.958	23.063		
Age (pedestrian)	22.2	< 0.001	2.142	1.491	3.078		
exp(β_0)			0.131				
MAIS 3+ (18+)						246	265
Weight vehicle	34.1	0.012	1.435	1.084	1.899		
Impact speed	28.9	< 0.001	17.853	8.945	35.632		
exp(β_0)			0.042				
MAIS 5+ (18+)						161	172
Impact speed	28.9	< 0.001	10.708	5.536	20.711		
Front-top transition point (height)	15.6	0.002	1.882	1.266	2.795		
exp(β_0)			0.007				

In addition to collision speed, physiological characteristics of the pedestrian (body height, age, sex, etc.), geometric characteristics of the vehicle, and various ratios linking these two factor groups were shown not only to be significant predictors of injury severity in univariate analysis, but also to be independent predictors in multivariate models. Thus, the hypothesis that a change in vehicle geometry could have an effect on injury severity is confirmed by this analysis even after controlling collision speed.

IN-SAMPLE PREDICTIVE ACCURACY

The in-sample predictive accuracy is quantified by ROC analysis. Table 7 summarizes the area under the ROC curves for the models of Table 4 and Table 5. Using this criterion, the in-sample predictive performance is remarkably high (ROC-AUC 0.879 to 0.943 for multivariate models compared to 0.880 to 0.899 for univariate models). The corresponding models containing only speed have lower in-sample performance.

Table 7. In-sample Predictive Accuracy of the ISS Models and Fatalities (age group 18+)

Model	ROC-AUC	Confidence interval		Model parameters (k)
ISS 9+ (age 18+)	0.879	0.840	0.918	3
ISS 9+ (age 18+) speed	0.880	0.837	0.923	1
ISS 16+ (age 18+)	0.908	0.875	0.942	3
ISS 16+ (age 18+) speed	0.890	0.850	0.930	1
ISS 25+ (age 18+)	0.943	0.915	0.970	5
ISS 25+ (age 18+) speed	0.899	0.854	0.944	1
Fatalities (age 18+)	0.936	0.903	0.969	5
Fatalities (age 18+) speed	0.885	0.832	0.938	1

Table 8 summarizes the ROC-AUC for the models of Table 6. The predictive performance is also high (ROC-AUC 0.830 to 0.904). Comparing MAIS and ISS levels, using their verbal description given above, i.e. at least moderate / serious / critical injury, it becomes obvious that MAIS based models have on average less predictive power (ROC-AUC 0.04 to 0.05 lower compared to the corresponding ISS model). It can be assumed that this effect is caused by the predominant ability of ISS to describe the overall injury severity of a person as explained above.

Table 8. In-sample Predictive Accuracy of the Models Using MAIS (age 18+)

Model	ROC-AUC	Confidence interval		Model parameters (k)
MAIS 2+ (age 18+)	0.830	0.784	0.876	2
MAIS 3+ (age 18+)	0.869	0.828	0.911	2
MAIS 5+ (age 18+)	0.904	0.860	0.947	2

VALIDATION OF OUT-OF-SAMPLE PREDICTIVE ACCURACY

The in-sample predictive accuracy presented above (see Table 7) is compared in Table 9 with the estimated out-of-sample accuracy estimated by the 10-fold cross validation. The optimism, defined as difference between ROC-AUC of the full-data model and the mean of the 10 cross-validation models, is relatively small; the out-of-sample predictive quality is still high. Each model is compared with a univariate model including only impact speed which represents the state of the art [Hannawald, 2004]. It is generally known and also becomes evident in this study that more model variables lead to more optimism.

P-Values in Table 9 refer to the hypothesis of improved ROC-AUC in the cross-validated multivariate models compared to the cross-validated speed-only models. The hypothesis cannot be accepted due to non significant differences in the mean. Considering the standard deviation (STD) as well as the standard error of the mean (SE), more data would be required to assess the predictive improvement of the multivariate model compared to the speed-only model. The STD as well as the SE are suspected to decrease with higher case numbers and become smaller than the optimism.

Table 9. Out-of-sample Predictive Accuracy of the Models (age group 18+)

Model	Full-data model ROC-AUC	Cross-validation ROC-AUC	p-Value	STD	SE	Optimism
ISS 9+ (age 18+)	0.879	0.863	0.743	0.067	0.022	0.016
ISS 9+ (age 18+) speed	0.862	0.867		0.058	0.019	-0.005
ISS 16+ (age 18+)	0.908	0.883	0.467	0.059	0.020	0.025
ISS 16+ (age 18+) speed	0.890	0.896		0.061	0.020	-0.006
ISS 25+ (age 18+)	0.943	0.919	0.111	0.059	0.020	0.024
ISS 25+ (age 18+) speed	0.899	0.897		0.094	0.031	0.002
Fatalities (age 18+)	0.936	0.891	0.900	0.070	0.023	0.045
Fatalities (age 18+) speed	0.885	0.893		0.057	0.019	-0.008

The 10-fold cross validation answers the question of out-of sample predictive accuracy compared to in-sample predictive accuracy of the full-data models. Another important issue is the stability of the variables used in each full-data model. Table 10 gives the number of cross validation models in which the variables of the full-data models of Table 4 and Table 5 are included. The stability of a variable rises with the number of cross-validation models in which it occurs. E.g., for ISS9+, it is obvious that impact speed and age of the pedestrian are included in every model, whereas the distance to the base of the windshield (measured as wrapping distance) is only included in two models. However one has to be careful to associate a low occurrence in cross-validation models automatically with instability of the variable. Sometimes different variables are included in the models. Suspected reason for this are few case numbers as well as collinearity. Especially geometric values of the front end are suspected to have some correlation among one another and therefore can be surrogates (for example ground to forward hood opening (wrap) and ground to front / top transition point).

The cross-validation for MAIS based models are not displayed as they do not enhance knowledge on the question whether to use ISS or MAIS. The effects described above for ISS also apply to MAIS based models.

Table 10. Stability of Variables in the Models

Full-data model / variables	variables of full-data models	pedestrian characteristics	vehicle characteristics
ISS 9+ (age 18+)		10	8
Impact speed	10		
Age (pedestrian)	10		
Ground to base of windshield (wrap)	2		
ISS 16+ (age 18+)		10	10
Impact speed	10		
Age (pedestrian)	10		
Forward hood opening (height)	5		
ISS 25+ (age 18+)		10	10
Impact speed	10		
Age (pedestrian)	10		
Forward hood opening (height)	6		
Weight vehicle	8		
Bumper leading angle	9		
Fatalities (age 18+)		10	8
Impact speed	6		
Kinetic energy	6		
Body mass index	10		
Age (pedestrian)	10		
Ratio height hip / front-top transition	3		

The methodology for assessing the general power of the models is presented. It shows that more explanatory variables lead to more optimism; however, the number of cases available does not allow a clear statement whether the multivariate models have or do not have more predictive power than univariate models including only impact speed.

DISCUSSION

As stated above, one focus of this paper is to suggest an approach toward comparison of active and passive safety measures with regard to their potential effect on injury severity. The basic metric is the risk reduction of injury probability and mortality, respectively. By taking the scaling factors and adjusted odds ratios of multivariate analysis into account, it is possible to perform quantitative comparisons of relevant influencing factors regarding their risk reduction potential, i.e., to estimate the impact of changes in these influencing factors on injury severity probabilities.

In order to illustrate how this analysis might be utilized to compare potential benefits of different candidate pedestrian protection approaches, we consider a benchmark computation for a prototype preventive pedestrian system using sensor-triggered emergency braking. As a benchmark we define a theoretical emergency braking system which is capable of reduing the impact speed by 15 km/h in a typical accident. (This reduction corresponds to an emergency braking system providing 8 m/s^2 deceleration for about a half of a second prior to impact, which appears feasible in many accident scenarios using currently available technology.) Table 11 gives the equivalent reduction of injury risk (age group 18+) for several severity levels quantified by ISS as well as for fatalities using Eq. (1) and the models given in Tables 4-5.

Table 11. Reduction of injury probability due to a 15 km/h reduction in impact speed (age group 18+).

Severity level	ISS9+	ISS16+	ISS25+	Fatal
Reduction of probability	25%	30%	31%	6%

Using these values as a benchmark, the improvements due to active safety can be compared to other measures. A reduction in injury probability quantified by ISS means that the measure evaluated effectively reduces injury severity concerning the severest injuries in up to three body regions.

The models for the prediction of injury probability of the pedestrian presented here can be compared to existing approaches. Note, that to this end a variation in impact speed is considered while all other factors remain constant. Eq. (3) of [Rosen, 2009b] gives anfatality probability of 6.0% for 50 km/h and 0.9% for 30 km/h, respectively. [Peden, M.; et al., 2004] cites several other studies concerning fatality probability. On the one hand, the probability at 50 km/h is eight times greater than at 30 km/h and on the other hand the probability at 30 km/h equals 10% and for 45 km/h is less than 50%. Using Eq. (1) and the values of Table 4 we computed the following probabilities: 3.5% (30 km/h), 14.9% (45 km/h), and 22.0% (50 km/h). The finding that the probability at 50 km/h is nearly eight times greater than at 30 km/h can be confirmed with minor deviations both from the results presented and from [Rosen, 2009b]. However, the absolute levels differ. The suspected reason for this are the different data sets used. [Rosen, 2009b] used the German In-depth Accident Study (GIDAS) which contains accidents from 1997 onwards sampled in Germany. The models presented here are based on PCDS.

The multivariate models are expressed in a form suitable for interfacing with large-scale statistical simulations. Those can quantify the benefits of preventive pedestrian safety systems which are designed to mitigate or avoid pedestrian collisions with passenger cars. A comparison of resulting injury severity due to measure of active safety is the first step to a methodology capable of comparing the efficacy of active and passive safety directly.

CONCLUSIONS

A systematic consideration of statistical results concerning injury probabilities in collisions provides a metric for virtually evaluating and comparing safety benefits of potential strategies for pedestrian protection, such as structural changes, passive devices, and active safety. The methodology used here for the US Pedestrian Crash Data Study (PCDS) could be applied in principle to any data set containing similar factors and outcome variables. The overall injury risk reduction due to a change in impact speed can be directly compared to the risk reduction due to proposed passive safety strategies.

In this paper multivariate injury (MAIS, ISS, NISS) and fatality models have been estimated from PCDS data. Several physiological as well as vehicle parameters are independent predictors in addition to age of the pedestrian and impact speed. The univariate results provide insights into the interaction of profile parameters of the vehicle on injury and fatality.

The current standard in predicting injury severity of the pedestrian can be improved. The high in-sample and out-of-sample predictive performance of the models due to additional factors was quantified by ROC analysis and 10-fold cross validation. We encourage the use of ISS instead of MAIS, as the overall injury level of the pedestrian is represented with higher accuracy, which becomes also obvious in the improved predictive power of the resulting multivariate models. As NISS and ISS nearly have identical values in the data set used, the question which is preferable cannot be answered by this study.

A similar investigation will be carried out using the German In-Depth Accident Statistics (GIDAS). Detailed information in vehicle characteristics has still to be created there. The issue of using NISS instead of ISS will also be considered again. A comparison of the two data sets as well as an application of the methodology presented to different body regions [Dahdah, 2008] is also planned.

Establishing priorities in approaches to pedestrian safety is a matter of public debate and policy. Since the shared objective of all stakeholders should be the optimal protection of pedestrians, rather than the advancement of a particular technology for its own sake, it seems appropriate to discuss proposed passive and active safety approaches using a common metric based on injury reduction. The methodology discussed in this paper is intended as a contribution to the effort toward defining such a common metric.

FUTURE NEEDS

PCDS represents a very good in-depth accident data source regarding pedestrian / vehicle collisions in the USA. However, when interpreting the results one must be aware that the data were collected in the 1990s and therefore may be outdated for some research questions. One has to take into consideration that the fleet represented in PCDS may have changed regarding geometry, stiffness, and other relevant characteristics regarding pedestrian injury causation. These limitations can only be overcome by continuing efforts in data collection. As detailed statistical analysis as well as construction of sub-samples requires sufficient case numbers, a larger database would also be beneficial. Considering the high stated priority of pedestrian protection among European agencies and in the international safety community, as well as the resources devoted to theoretical and political discussion of the subject, it is surprising (to say the least) that empirical data resources are so scarce. There is a clear need for additional large-scale, high-quality, up-to-date, internationally coordinated collection of in-depth accident data on vehicle / pedestrian collisions

ACKNOWLEDGEMENTS

TH gratefully acknowledges a fellowship of the German Academic Exchange Service.

REFERENCES

[AAAM, 1990] AAAM. 1990. The Abbreviated Injury Scale (1990 revision). *Association for the Advancement of Automotive Medicine*.

[Baker, 1976] Baker, S. P.; O'Neill, B. 1976. The Injury Severity Score: An Update. *The Journal of Trauma*, **16**(11), 882–885.

[Baker, 1974] Baker, S. P.; O'Neill, B.; Haddon W. Jr.; Long W. B. 1974. The Injury Severity Score: A Method for Describing Patients With Multiple Injuries and Evaluating Emergency Care. *The Journal of Trauma*, **14**(3), 187–196.

[Dahdah, 2008] Dahdah, S. 2008. *Modeling an infrastructure safety rating for vulnerable road users in developing countries*. Dissertation, The George Washington University.

[Ebner, 2010] Ebner, A.; Samaha, R. R.; Scullion P.; Helmer T. 2010. Methodology for the development and evaluation of active safety systems using reference scenarios. *17th ITS World Congress*, (in press).

[Friesen, 2003] Friesen, F. A. 2003. *Motorhaubenaufstellung für den Fußgängerschutz*. Forschungsgesellschaft Kraftfahrwesen Aachen mbH (fka).

[Gabler, 2005] Gabler, C.; Digges, K.; Fildes B.; Sparke L. 2005. Side Impact Injury Risk for Belted Far Side Passenger Vehicle Occupants. *SAE transactions*, **114**(6), 34–42.

[Hakkert, 2007]Hakkert, A. S.; Gitelman, V.; Vis M. A. 2007. *Road Safety Performance Indicators: Theory*. Tech. rept. Deliverable D3.6 of the EU FP6 project SafetyNet.

[Hannawald, 2004] Hannawald, L.; Kauer, F. 2004. *Equal Effectiveness Study On Pedestrian Protection*. Dresden: Technische Universität Dresden.

[Helmer, 2010] Helmer, T.; Scullion, P.; Samaha R. R.; Ebner A.; Kates R. 2010. Predicting the injury severity of pedestrians in frontal car crashes based on empirical, in-depth accident data. *17th ITS World Congress*, (in press).

[Henary, 2003] Henary, B. Y.; Crandall, J.; Bhalla K.; Mock C. N.; Roudsari B. S. 2003. Child and adult pedestrian impact: The influence of vehicle type on injury severity. *In: 47th Annual Conference of the Association for the Advancement of Automotive Medicine*.

[Inomata, 2009] Inomata, Y.; Iwai, N.; Maeda Y.; Kobayashi S.; Okuyama H.; Takahashi N. 2009. Development Of The Pop-up Engine Hood For Pedestrian Head Protection. *In: 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV 2009)*.

[Isenberg, 1998] Isenberg, R. A.; Chidester, A. B.; Mavros S. 1998. Update On The Pedestrian Crash Data Study. *16th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*.

[Jarrett, 1998] Jarrett, K. J.; Saul, R. A. 1998. Pedestrian Injury-Analysis Of The PCDS Field Collision Data. *In: 16th International Technical Conference on the Enhanced Safety of Vehicles (ESV 1998)*.

[Kerrigan, 2008] Kerrigan, J. R.; Rudd, R.; Subit D.; Untaroiu C. D.; Crandall J. R. 2008. Pedestrian Lower Extremity Response and Injury: A Small Sedan vs. A Large SUV. *SAE World Congress*.

[Kühn, 2005] Kühn, M.; Fröming, R.; Schindler V. 2005. *Fußgängerschutz - Unfallgeschehen, Fahrzeuggestaltung, Testverfahren*. Berlin: Springer.

[Lefler, 2001] Lefler, D. E.; Gabler, H. C. 2001. The emerging threat of light truck impacts with pedestrians.

[Lefler, 2004] Lefler, D. E.; Gabler, H. C. 2004. The fatality and injury risk of light truck impacts with pedestrians in the United States. *Accident Analysis and Prevention*, **36**, 295–304.

[Longhitano, 2005] Longhitano, D.; Henary, B.; Bhalla K.; Ivarsson J.; Crandall J. 2005. Influence of Vehicle Body Type on Pedestrian Injury Distribution. *2005 SAE World Congress*.

[Nations, 2009] Nations, United. 2009. *Global Technical Rule No. 9 - Pedestrian Safety*. United Nations.

[NCAP, 2010] NCAP, Euro. 2010. *European New Car Assessment Programme (Euro NCAP) Pedestrian Testing Protocol*. www.euroncap.com.

[Nogueira, 2008] Nogueira, L. S.; Domingues, C. A.; Campos M. A.; Sousa R. M. C. 2008. Ten years of New Injury Severity Score (NISS): Is it a possible change? *Rev Latino-am Enfermagem*, **16**(2), 314–319.

[Ogden, 2004] Ogden, C. L.; Fryar, C. D.; Carroll M. D.; Flegal K. M. 2004. *Mean body weight, height, and body mass index, United States1960–2002*. Advance data from vital and health statistics 347. National Center for Health Statistics, Hyattsville, Maryland.

[Osler, 1997] Osler, T.; Baker, S. P.; Long W. 1997. A Modification of the Injury Severity Score That Both Improves Accuracy and Simplifies Scoring. *The Journal of Trauma: Injury, Infection, and Critical Care*, **43**(6), 992–926.

[Peden, M.; et al., 2004] Peden, M.; et al. 2004. *World report on road traffic injury prevention*. World Health Organization.

[Rosen, 2009a] Rosen, E.; Källhammer, J.-E.; Eriksson D.; Nentwich M.; Frederiksson R.; Smith K. 2009a. Pedestrian Injury Mitigation By Autonomous Braking. *In: 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV 2009)*.

[Rosen, 2009b] Rosen, E.; Sander, U. 2009b. Pedestrian fatality risk as a function of car impact speed. *Accident Analysis and Prevention*, **41**, 536–542.

[Roudsari, 2005] Roudsari, B. S.; Mock, C. N.; Kaufman R. 2005. An Evaluation of the Association Between Vehicle Type and the Source and Severity of Pedestrian Injuries. *Traffic Injury Prevention*, **6**, 185–192.

[Scullion, 2010] Scullion, P.; Samaha, R. R.; Helmer T.; Ebner A. 2010. US Pedestrian Crashes: Current Trends. *In: Canadian Multidisciplinary Road Safety Conference (in press)*.

[Simms, 2006] Simms, C. K.; Wood, D. P. 2006. Pedestrian risk from cars and sport utility vehicles – a comparative analytical study. *Pages 1085–1100 of: Proceedings of the Institution of Mechanical Engineers*, vol. 220.

[States, 1980] States, J. D.; Huelke, D. F. 1980. The Abbreviated Injury Scale. 1980 Revision. American Association for Automotive Medicine (AAAM).

[States, 1969] States, J. D. 1969. Abbreviated and the Comprehensive Research Injury Scales. *Pages 282–294 of: Proceedings of Thirteenth Stapp Car Crash Conference, December 2-4, 1969*, vol. 13. Society of Automotive Engineers, New York.

[Stevenson, 2001] Stevenson, M.; Segui-Gomez, M.; Lescohier I.; Di Scala C.; McDonald-Smith G. 2001. An overview of the injury severity score and the new injury severity score. *Injury Prevention*, **7**, 10–13.

[UMTRI, 2005]UMTRI. 2005. *1994-1998 NASS Pedestrian Crash Data Study (PCDS) Codebook. Version 03Mar01*. UMTRI Transportation Data Center.

[Union, 2009] Union, Europäische. 2009. *Verordnung (EG) Nr. 78/2009*. Amtsblatt der Europäischen Union.

[Wisselmann, 2009] Wisselmann, D.; Gresser, K.; Hopstock M.; Huber W. 2009. Präventiver statt passiver Fußgängerschutz. *Pages 60–76 of:* Gesamtzentrum für Verkehr Braunschweig e.V. (ed), *AAET 2009 - Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel*.

APPENDIX

Table 12. Continuous variables used.

Variable name	Unit	N	Mean	SD
Pedestrian:				
Age (pedestrian)	a	449	35.9	22.2
Weight (ped., imputed)	kg	450	66.6	22.6
Body height (ped., imputed)	cm	450	162.8	19.5
Body mass index	kg/m^2	450	24.4	5.7
Knee height (ped., imputed)	cm	449	46.7	5.9
Hip height (ped., imputed)	cm	449	88.2	11.1
Shoulder height (ped., imputed)	cm	449	134.2	16.4
Vehicle:				
Weight vehicle	kg	450	141.5	34.1
Hood length	cm	450	102.7	19.5
Front bumper (bottom height)	cm	450	38.7	6.5
Front bumper (top height)	cm	450	54.8	6.6
Forward hood opening (height)	cm	450	76.1	17.0
Front bumper lead	cm	449	9.3	3.0
Ground to forward hood opening (wrap)	cm	450	80.3	14.4
Front-top transition point (height)	cm	450	85.9	15.6
Ground to rear hood opening (wrap)	cm	450	184.8	22.2
Ground to base of windshield (wrap)	cm	450	194.5	22.6
Ground to top of windshield (wrap)	cm	450	271.8	21.1
Bumper leading angle	°	449	61.9	15.4
Collision:				
Impact speed	km/h	376	28.9	20.8
Impact speed (squared)	$(km/h)^2$	376	-	-
Kinetic energy	kJ	376	70.0	113.1
Ratios:				
Ratio height knee / bumper bottom	-	449	1.2	0.3
Ratio height knee / bumper top	-	449	0.9	0.1
Ratio height hip / front-top transition	-	449	1.1	0.2
Ratio height hip / forward hood opening	-	449	1.2	0.3
Ratio height hip / ground forw. hood op.	-	449	1.1	0.2
Ratio height shoulder / ground rear hood op.	-	449	0.7	0.1
Ratio height shoulder / ground front hood op.	-	449	1.7	0.4
Ratio body height / front-top transition	-	450	2.0	0.4
Ratio body height / ground rear hood opening	-	450	0.9	0.2
Ratio body height / ground front hood opening	-	450	2.1	0.4
Ratio body height / ground base windshield	-	450	0.8	0.1
Ratio body height / ground top windshield	-	450	0.6	0.1
Ratio body height / hood length	-	450	1.7	0.5

Table 13. Recoding of non-continuous variables (part 1).

Original variable	N	Original coding	New variable	N	New coding
Sex	**376**		**Sex**	**376**	
Male	191	1	male (1)	191	1
Female - not reported pregnant	182	2	female (2 - 6)	185	2
Female - pregnant - 1st trimester (1st-3rd month)	2	3			
Female - pregnant - 2nd trimester (4th-6th month)	1	4	**Pregnancy**	**376**	
Female - pregnant - 3rd trimester (7th-9th month)	0	5	not pregnant (1, 2)	373	0
Female - pregnant - term unknown	0	6	1st trimester (3)	2	1
Unknown	0	9	2nd trimester (4)	1	2
Action of the pedestrian	**376**		**Direction ped.: away from lane**	**376**	
Stopped	10	0	walking away from lane (6)	2	1
Crossing road - straight	278	1	not walking away from lane (rest)	374	0
Crossing road - diagonally	58	2			
Moving in road - with traffic	10	3	**Direction ped.: with; against traffic**	**376**	
Moving in road - against traffic	2	4	with; against traffic (3, 4, 7, 9)	15	1
Off road - approaching road	0	5	not with; against traffic (rest)	361	0
Off road - going away from road	2	6			
Off road - moving parallel	3	7	**Direction ped.: towards lane; crossing**	**376**	
Off road - crossing driveway	10	8	towards lane; crossing (1, 2, 5, 8)	346	1
Off road - moving along driveway	0	9	not towards lane; crossing (rest)	30	0
Other	1	98			
Unknown	2	99			
Motion of the pedestrian	**373**		**Walking: speed (2, 3)**	**376**	
Not moving	12	0	low speed (2, 3)	201	0
Walking slowly	182	1	high speed (rest)	175	1
Walking rapidly	36	2			
Running or jogging	139	3	**Walking: danger**	**376**	
Hopping	0	4	low danger (rest)	186	0
Skipping	0	5	high danger (0, 2, 3, 7)	190	1
Jumping	0	6			
Falling/stumbling or rising	3	7			
Other	1	8			
Unknown	3	9			
Attempted avoidance maneuver of the car	**376**		**Avoidance: Braking without lockup**	**376**	
No avoidance actions	144	1	not braking without lockup (rest)	221	0
Braking (no lockup)	104	2	braking without lockup (2, 8, 9)	155	1
Braking (lockup)	65	3			
Braking (lockup unknown)	3	4	**Avoidance: Braking with lockup**	**376**	
Releasing brakes	0	5	not braking with lockup (rest)	311	0
Steering left	4	6	braking with lockup (3)	65	1
Steering right	2	7			
Braking and steering left	29	8	**Avoidance: Releasing brakes**	**376**	
Braking and steering right	19	9	not releasing brakes (rest)	376	0
Accelerating	1	10	releasing brakes (5)	0	1
Accelerating and steering left	0	11			
Accelerating and steering right	0	12	**Avoidance: Accelerating**	**376**	
Other action	0	98	not accelerating (rest)	375	0
Unknown	5	99	accelerating (10 - 12)	1	1
			Avoidance: Steering left	**376**	
			not steering left (rest)	343	0
			steering left (6, 11, 29)	33	1
			Avoidance: Steering right	**376**	
			not steering right (rest)	355	0
			steering right (7, 9, 12)	21	1
			Avoidance: Steering	**376**	
			not steering (rest)	322	0
			steering (6 - 9, 11, 12)	54	1

Table 14. Recoding of non-continuous variables (part 2).

Original variable	N	Original coding	New variable	N	New coding
Driver attention	**376**		**Driver's attention**	**376**	
Full attention to driving	310	1	driver not distracted (1)	310	0
Distracted by other occupant	6	2	driver distracted (rest)	66	1
Distracted by moving object in vehicle	1	3			
Distracted by outside person, object or event	21	4			
Talking on cellular phone/CB radio	2	5			
Sleeping or dozing while driving	3	6			
Other	21	8			
Unknown	12	9			
Pre-event movement of the car	**375**		**Pre-event movement car: complexity**	**376**	
No driver present	232	0	low complexity (0 - 4)	246	0
Going straight	9	1	high complexity (rest)	130	1
Slowing or stopping in traffic lane	3	2			
Starting in traffic lane	2	3			
Stopped in traffic lane	7	4			
Passing or overtaking another vehicle	34	5			
Disabled or parked in travel lane	0	6			
Leaving a parking position	0	7			
Entering a parking position	0	8			
Turning right	75	9			
Turning left	1	10			
Making U-turn	3	11			
Backing up (other than for parking position)	0	12			
Negotiating a curve	5	13			
Changing lanes	1	14			
Merging	2	15			
Successful avoidance maneuver to a previous critical event	1	16			
Other	1	97			
Unknown	1	99			
Body type	**376**		**Body type (only automobiles)**	**376**	
Automobiles			automobiles (1-6)	246	1
	5	1	utility vehicles, van based light trucks,	130	0
Convertible (excludes sun roof, t-bar)			conventional light trucks (rest)		
2-door sedan, hardtop, coupe	51	2			
3-door/2-door hatchback	34	3			
4-door sedan, hardtop	142	4			
5-door/4-door hatchback	9	5			
Station Wagon (excluding van and truck based)	5	6			
Utility Vehicles (< 4,536 kgs GVWR)					
Compact utility (includes Jeep CJ-2 thru Jeep CJ-7, Renegade, Wrangler, Navajo, Bronco, Thing)	20	14			
Large utility (includes Jeep Cherokee, Ram Charger, Trailduster, fullsize Blazer, Scout)	2	15			
Utility station wagon (Chevy Suburban,GMC Suburban, Travelall,Grand Wagoneer,includes suburban limousine)	4	16			
Van Based Light Trucks (< 4,536 kgs GVWR)					
Minivan (Astro, Caravan, Plymouth Vista, Aerostar, Safari, et	41	20			
Large van (Sportsman, Royal, Ram, Chevy Van, Voyager-Pre83, Econoline, etc.)	14	21			
Step van or walk-in van (< 4,536 kgs GVWR)	1	22			
Lt Conventional Truck (P.U. style cab, < 4,536 kgs. GVWR)					
Compact pickup (Ram 50, Dakota, Arrow Pickup, Ranger, Toyota Pickup, etc.)	17	30			
Large pickup (Jeep Pickup, Comanche, Silverado, Sierra, etc	31	31			
Unknown body type	0	99			