Influence of impact speed estimation errors on pedestrian fatality risk curves

E Rosén, U Sander

Autoliv Research, Wallentinsvägen 22, 447 83, Vårgårda, Sweden.

Abstract – Recent findings from real-world accident data have shown that fatality risks for pedestrians are substantially lower than generally reported in the traffic safety literature. One of the keys to this insight has been the large and random sample of car-to-pedestrian crashes available in the German In-Depth Accident Study (GIDAS). Another key factor has been the proper use of weight factors in order to adjust for outcome-based sampling bias in the accident data. However, a third factor, a priori of unknown importance, has not yet been properly analysed. This is the influence of errors in impact speed estimation. In this study, we derived a statistical model of the impact speed errors for pedestrian accidents present in the GIDAS database. The error model was then applied to investigate the effect of the estimation error on the pedestrian fatality risk as a function of car impact speed. To this end, we applied a method known as the SIMulation-EXtrapolation (SIMEX) method. It was found that the risk curve is fairly tolerant to some amount of random measurement error, but that it does become flattened. It is therefore important that the accident investigations and reconstructions are of high quality to assure that systematic errors are minimised and that the random errors are under control.

Keywords: Pedestrian; fatality risk; logistic regression; impact speed; estimation error, SIMEX

1. INTRODUCTION

In a previous study of the real-world accident data available in the GIDAS database we pointed out that pedestrian fatality risks have been largely exaggerated during the past decades [1]. A sequel review of the available literature on the subject clearly showed that previous results were generally based on data with a substantial bias towards severe and fatal accidents, thus explaining the excessively high fatality risks [2]. To exemplify, previous results had demonstrated that the risk of death for a pedestrian struck by a car at 50 km/h was between 40% and 90% [3–5], whereas the recent findings indicate a fatality risk of approximately 10% [1, 6, 7]. However, it is important to note that there is still a five-fold increase in fatality risk when impact speed increases from 30 km/h to 50 km/h. Hence, these new findings still supports the previous conclusions that speed should be as low as possible where car-to-pedestrian crashes are at risk. While the previous results claimed that impact speeds exceeding 60 km/h were practically un-survivable, the new findings shows that there is a large benefit in braking a car from, e.g., 80 km/h to 60 km/h. This could be achieved by manually or autonomously activated brake assist systems. The former have existed for several years, while the latter have reached the market this year.

In both Refs. [2] and [6], the potential influence of estimation errors in impact speeds on fatality risk curves was pointed out. In this study, we aimed to quantify this influence by deriving an error model applicable to the car-to-pedestrian crashes in GIDAS and then applying the SIMulation-EXtrapolation (SIMEX) method [8] to adjust for the estimation errors.

3. METHODS

3.1 Fatality risk curve

The starting point of this study was the fatality risk curve of Ref. [1], which was derived from a sample of pedestrians aged 15 years and older that were struck by the front of a passenger car. Pedestrians that were lying on the ground prior to impact were excluded. This yielded a sample of 490 pedestrian accidents, including 36 fatalities, from the GIDAS database from 1999 to 2007. Weight factors were derived by comparing to national statistics in order to adjust for sampling bias [9]. Logistic regression analysis was applied to the weighted sample in order to derive an analytical expression for the pedestrian fatality risk as a function of impact speed. The fatality risk (probability of death), $P(v)$, was then assumed to have the following form (logistic regression)

$$P(v) = \frac{1}{1 + e^{-a-bv}} \,, \qquad [1]$$

where v is the impact speed in km/h and a, b, two parameters to be estimated by the method of maximum likelihood [10, 11]. In Ref. [1], it was found that $a = -6.9$ and $b = 0.090$ (the risk curve with 95% confidence intervals is reproduced in Figures 2 and 4). This was the starting point of the present study.

3.2 Estimation of impact speed in GIDAS

GIDAS pedestrian accident reconstructions are primarily based on information collected at the accident site during the on-scene investigations. Securing of transient evidence, e.g., collision point, pedestrian and car end-positions, and brake marks, is essential for the quality of the reconstruction. Furthermore, on-scene investigations facilitate interviews with the driver, pedestrian and other eye-witnesses. The choice of reconstruction method depends on the availability and reliability of the collected data [12].

In general, all reconstruction methods are based on the physical principles of conservation of momentum and energy (work-energy principle). An approximate car impact speed can be calculated from the energy dissipated by braking from the collision point to the car end-position. This estimate can be further improved by taking the pedestrian walking velocity together with the mass ratio of the car and pedestrian into account. When brake marks are absent, the pedestrian throw distance, defined as the distance between the collision point and the pedestrian end-position, can also be used to estimate the impact speed of the car. The throw distance can be divided into a flying phase and a gliding / rolling phase of the pedestrian over the ground. The length of each phase is influenced by the car impact speed, the car front-end shape, and if the car was braked or not at the time of collision. Therefore the usage of a validated multi-body model of the pedestrian and a 3D car shape in a simulation environment, e.g. PC Crash, is recommendable. Nevertheless, some generic relations between impact speed, throw distance, car front-end shape, and contact offset of the pedestrian along the front-end have been published based on experimental data [13–17]. It is noted that most of these experimental data were based on the front-end shapes of older car generations. Furthermore, recent research has shown that using the offset between pedestrian stature and the wrap around distance to the head contact point on the car (pedestrian gliding distance) does not provide reliable impact speeds [18]. The upper bound method combines the above mentioned methods to lower the uncertainty and also takes into consideration special restrictions and speed limitations due to environmental factors. At a last instance, witness statements can be used to validate data for the accident reconstruction.

The following information was based on private communication with the persons responsible for reconstructions at the GIDAS centres in Dresden and Hanover: About 85% of the pedestrian accident reconstructions in GIDAS are based on the car brake distance and estimated deceleration (considering the coefficient of friction). Typically, for urban accidents with lower impact speeds, the pedestrian throw distance is either unknown or small compared to the size of the pedestrian, which makes a definition of the throw distance nearly impossible. Generally the reconstructions are conducted with use of PC Crash; most often with a rigid pedestrian model (DAT-file) and less often with a multi-body model. In about 10% of the accidents, the collision point and end-positions of pedestrian and car are available and reliable. The typical scenario for this is an accident in a rural region with higher car impact speed. Besides the car kinematics, also the pedestrian kinematics and the interaction with the car are analysed by using a multi-body model within PC Crash. In less than 5% of the accidents, the on-scene information is not sufficient for a detailed accident reconstruction. Examples are “hit and run” accidents or crashes with minor injury outcome, so that the accident site is already cleaned up at the arrival of the GIDAS investigation team. In these cases, impact speed estimations are primarily based on witness statements.

3.3 Estimation error in GIDAS

The derivation of risk curves is sensitive to systematic errors in the impact speed estimations [19]. A systematic error towards higher or lower impact speeds would inevitably shift the risk curve to the right or left respectively. In order to run a small test on the plausibility of the GIDAS reconstructions, the median impact speed for fatalities in GIDAS has been compared to two other pedestrian real-world accident investigations [1]. In this comparison, it was found that the median was 57 km/h in the GIDAS database and between 50 km/h and 60 km/h in both an Australian accident database ($N_{\text{fatal}} = 181$) from 1981 and British in-depth data from the 1960s and 70s ($N_{\text{fatal}} = 81$) [20, 21]. Hence, the median impact speeds for the fatally wounded pedestrians in these three large real-world accident studies are in accord with each other. This provides an indication that any systematic error in impact speed was small in these three databases, unless they were all biased the same way.

To investigate the random error in the GIDAS speed estimations, we started by considering the case when brake marks from permanent or temporarily locked-up wheels are identifiable. The investigators should then strive to find the exact spot on the ground where the pedestrian was struck by the car and the car end-position. The distance between those two points, along the lines of the brake marks, is the relevant brake distance, *d*. Let us denote the estimated brake distance by d' and define the estimation error as $\Delta d \equiv d' - d$. Furthermore, denote the brake acceleration by *a*, the estimated brake acceleration by a', and then define the estimation error as $\Delta a \equiv a' - a$. (When skid marks are present, $a = \mu g$, where μ is the dynamic coefficient of friction and $g = 9.82\, m/s^2$ is the gravitational constant.) The true car impact speed is related to the brake distance and acceleration (deceleration) as $v = \sqrt{2ad}$ (note that the absolute value of *a* must be used). The estimated impact speed would be calculated as $v' = \sqrt{2a'd'}$. The estimation error of impact speed is defined as $\Delta v \equiv v' - v$. (Note that the impact speed needs to be slightly adjusted in order to take the mass ratio of the car and pedestrian into account. However, this step infers only minor errors to the calculations and is therefore neglected in this example.) A simple calculation now yields that

$$\frac{\Delta v}{v} = \frac{v'}{v} - 1 = \sqrt{\frac{1}{2ad} 2(a + \Delta a)(d + \Delta d)} - 1 = \sqrt{\left(1 + \frac{\Delta a}{a}\right)\left(1 + \frac{\Delta d}{d}\right)} - 1 \approx 0.5\left(\frac{\Delta a}{a} + \frac{\Delta d}{d}\right), \qquad [2]$$

where the last equality is approximately true when the errors in brake acceleration and distance are small, i.e., $\Delta a/a << 1$ and $\Delta d/d << 1$. (When the errors are not small, the last equality is a poor approximation and $\Delta v/v$ can instead be simulated using the next to last expression.) From equation [2], it follows that the variance of $\Delta v/v$ can be approximated by

$$\mathrm{Var}\left(\frac{\Delta v}{v}\right) \approx 0.25\left(\mathrm{Var}\left(\frac{\Delta a}{a}\right) + \mathrm{Var}\left(\frac{\Delta d}{d}\right)\right). \qquad [3]$$

It is plausible to assume that the estimated brake acceleration, a', and distance, d', are normally distributed around the true values of brake acceleration, *a*, and distance, *d*. Furthermore, the size of the error likely increases with the true values of *a* and *d*. An optimistic assumption is that the standard deviations of the errors Δa and Δd equal 10% of the true brake acceleration and distance. This means that the estimated value of *a* will fall within plus minus 10% of the true value in 67% of the cases and within plus minus 20% of the true value in 95% of the cases. Since the standard deviation is defined as the square root of the variance, this gives that $\mathrm{Var}(\Delta a/a) = \mathrm{Var}(\Delta d/d) = 0.01$. It then follows from equations [2] and [3] that Δv is normally distributed around the true value of impact speed with variance $\mathrm{Var}(\Delta v/v) = 0.005$. Hence, the standard deviation of Δv becomes $\sqrt{0.005} = 7\%$ of the true impact speed, *v*. On the other hand, a pessimistic assumption would be that the standard deviations of delta Δa and Δd are 20% of the true values of *a* and *d*. Equation [3] then implies that the standard deviation of Δv is 14% of the true impact speed. (In this case $\Delta a/a = \Delta d/d = 0.2$ which is not much smaller than 1. Hence, the final equality in equation [2] may be affected by higher order corrections. The interested reader can proceed by simulating $\Delta v/v$ using the next to last equality in equation [2] to find that it is close to normally distributed with a standard deviation close to 14%.) Hence, in cases where brake marks are identifiable, the GIDAS teams are likely to provide impact speed estimations which are normally distributed around the true impact speed with a standard deviation substantially less than 15% of the true impact speed.

When brake marks cannot be identified, the pedestrian injury pattern can be used to validate the car brake status at the initial pedestrian contact. The deceleration level is then set in accordance with environmental conditions and driver or witness statements. It is understood that the average error of the estimated impact speed is greater under these circumstances compared to when brake marks are identified, since both *a* and *d* will be more difficult to estimate. A realistic, but slightly pessimistic, assumption is that $\mathrm{Var}(\Delta a/a) = 0.09$ (i.e., the standard deviation $\sigma(\Delta a/a) = 30\%$) and $\mathrm{Var}(\Delta d/d) = 0.04$ (i.e., $\sigma(\Delta d/d) = 20\%$), which gives that $\mathrm{Var}(\Delta v/v) = 0.0325$ using equation [3]. Hence, the standard deviation of the estimation error is approximately 18% of the true impact speed.

The pedestrian throw distance, if available, can also be used to estimate the car impact speed. Generally the error in the glide / roll distance of the pedestrian is similar to the error of the car brake distance. However, the coefficient of friction between the pedestrian and the ground may be difficult to estimate, depending on the present conditions. If air resistance is neglected it can be assumed that the horizontal velocity at the start of the flying phase is the same at the end of the flying phase, which is the input velocity for the gliding / rolling phase. The throw distance is most sensitive to errors when the throw angle is close to 0 or 90 degree, which can be shown by a Taylor series expansion of the error term. The actual calculations are similar to the case when brake marks are present, since the same principle of conservation of energy is applied. Nevertheless, the total amount of random estimation error is greater when pedestrian throw distance is used instead of car brake marks. Applying equation [3] with the reasonable assumptions $\mathrm{Var}(\Delta a/a) = 0.09$ (i.e., $\sigma(\Delta a/a) = 30\%$) and $\mathrm{Var}(\Delta d/d) = 0.04$ (i.e., $\sigma(\Delta d/d) = 20\%$) gives $\mathrm{Var}(\Delta v/v) = 0.0325$, which means that the standard deviation of the estimation error is approximately 18% of the true impact speed. (Simulation using the next to last expression in equation [2] confirms these findings.)

When both car brake marks and the pedestrian throw distance are available, both methods can be applied with the aim to narrow the uncertainty of the estimated impact speed. Applying the upper bound method in these cases, the standard deviation of the estimation error is expected to be less than 10% of the true impact speed.

3.4 Error models

From the considerations of section 3.2 and 3.3, it follows that the different reconstruction methods used by GIDAS all provide estimated impact speeds, *w*, that have a random error that is approximately normally distributed around the true (unknown) impact speed, *v*. Furthermore, the standard deviation of the random error is likely to, at least approximately, increase linearly with impact speed. In mathematical terms this leads to a multiplicative error model, which can be expressed as

$$w = v(1+u), \qquad [4]$$

where *u* is a random error term that is normally distributed with mean=0 and constant variance Var_u (i.e. $u \sim N(0, \mathrm{Var}_u)$). The total estimation error is therefore $vu \sim N(0, v^2 \mathrm{Var}_u)$. Hence, the standard deviation of the total error, *vu*, is linearly proportional to the true impact speed, $\sigma(vu) = v\sqrt{\mathrm{Var}_u}$. From sections 3.2 and 3.3, Var_u depends on the choice of reconstruction method. However, on average it is likely that Var_u should be around $0.15^2 = 0.0225$, corresponding to a random error with a standard deviation equal to 15% of the true impact speed. We chose to proceed by using three different multiplicative error models in which the standard deviations were 10%, 20%, and 30% of the true impact speed, respectively. This corresponded to Var_u =0.01, 0.04, and 0.09, respectively. See Table 1. Error model 1 was likely too optimistic and error model 3 too pessimistic.

Note that the GIDAS database includes a variable named “VKPM”, which, according to the GIDAS code book, gives the tolerance of the impact speed estimate in km/h. This variable has been filled out by the Hanover centre since 1999 and by the Dresden centre since 2005. In total, 253 of the 490 cases analysed in Ref. [1] had a given tolerance. In order to use the coded tolerance in a statistical analysis, its exact meaning was needed. One could possibly interpret the tolerance as giving a 95% confidence interval for the true impact speed, which would mean that the given tolerance equalled 2σ. However, such a statistical definition was not available, which made it difficult to use this variable in the present study. In nearly 80% of the cases, the tolerance was ± 5 km/h or ± 10 km/h. In some cases, the tolerance equalled the estimated impact speed, e.g., the impact speed was given as 30 km/h and the tolerance as ± 30 km/h. This was likely due to miscoding. Removing these cases, it was found that the tolerance increased with impact speed up to about 15 km/h. Above 15 km/h, there was no association between the tolerance and the estimated impact speed. Hence, an additive error model would be the best choice at impact speeds above 15 km/h whereas a multiplicative model should be best at impact speeds up to 15 km/h. For that reason, the SIMEX analysis was carried out for three different additive models as well. In all three models, the estimation error was normally distributed around the true impact speed, and the standard deviation was 5 km/h, 10 km/h, and 15 km/h respectively, see Table 1.

To summarise, three models assumed a multiplicative error, which means that the standard deviation of the estimation error increased linearly with the true impact speed. The other three models assumed an additive error, which means that the standard deviation of the estimation error was constant for all impact speeds. For all models, the error was assumed to be normally distributed around the true impact speed. See Table 1.

Table 1. Summary of error models. Note that *v* is the true, unknown, impact speed.

Error model	*Standard deviation of random error, σ*
Multiplicative error model 1	0.1*v*
Multiplicative error model 2	0.2*v*
Multiplicative error model 3	0.3*v*
Additive error model 1	5 km/h
Additive error model 2	10 km/h
Additive error model 3	15 km/h

3.5 The SIMEX method

The 490 accidents analysed in Ref. [1] provided equally many paired observations denoted by (y_i, w_i), i=1,…, 490; where $y_i = 1$ if pedestrian i died, $y_i = 0$ if pedestrian i survived and w_i was the estimated car impact speed for the i:th accident. The true, unknown, impact speed is denoted by v_i. The basic regression model that one would have liked to analyse for the 490 cases is

$$\text{logit}(\Pr\ y = 1 \mid v) = a_v + b_v v\ , \qquad [5]$$

where a_v and b_v are the regression coefficients. It is convenient to introduce the vector $\theta_v = (a_v; b_v)$. Since only the estimated impact speeds, w_i, and not the true impact speeds, v_i, were available, the logistic regression model analysed in Ref. [1] was

$$\text{logit}(\Pr\ y = 1 \mid w) = a_w + b_w w\ , \qquad [6]$$

which resulted in $\theta_w = (-6.9; 0.090)$. In order to get an estimate of the "true" regression coefficients, θ_v, the basic idea of the SIMEX method is now to add random error to the estimated impact speeds, w_i, and study how the available regression coefficients, θ_w, changes. It then uses a simple trick to estimate the effect of removing random error from the estimated impact speeds, w_i.

For the multiplicative error models (see Table 1), the SIMEX analysis starts by introducing a new random variable $W_b(z)$ as

$$W_b(z) = w\left(1 + U_b\sqrt{z}\right), \qquad [7]$$

with $U_b \sim N(0, \text{Var}_u)$, $b = 1,...,50$, and z a real number that determines the size of the added error. In this study, the following values of z were studied, z=0, 0.5, 1.0, 1.5, 2.0. (These particular values of b and z are standard in the literature [8].) The $W_b(z)$ can be referred to as the remeasured impact speeds. Values of $W_b(z)$ for each of the 490 observations and for each value of b and z were simulated using the statistical software SAS version 9.1.3. Note that $W_b(z = 0) = w$ and that with v, u, and U_b independent for all values of b

$$\text{E}(W_b(z) \mid v) = v \qquad [8]$$

$$\text{Var}(W_b(z) \mid v) = v^2 \text{Var}_u \cdot (1 + z(1 + \text{Var}_u)). \qquad [9]$$

From equation [9], it follows that the variance of $W_b(z)$ increases with z. Furthermore, choosing z negative, the variance of $W_b(z)$ becomes smaller than the variance of w (note that Var(w) = $v^2\text{Var}_u$). Hence, this can be interpreted as removing random error from w. For the particular choice of $z = -1/(1 + \text{Var}_u)$, the variance of $W_b(z)$ equals 0, which is a key property for the remeasured data. This means that the variable $W_b(z = -1/(1 + \text{Var}_u))$ should be representative for the true impact speed v when used in a logistic regression analysis.

For the additive error models (see Table 1), the remeasured impact speeds are introduced as $W_b(z) = w + U_b\sqrt{z}$, with $U_b \sim N(0, \text{Var}_u)$ and b and z similar as for the multiplicative models. This gives that $\text{E}(W_b(z) \mid v) = v$ and $\text{Var}(W_b(z) \mid v) = v^2 \text{Var}_u \cdot (1 + z)$. Hence, the variance of the remeasured impact speeds equals 0 at $z = -1$.

The SIMEX method proceeds by analysing the following regression model:

$$\text{logit}\left(\Pr\, y = 1 \mid W_b(z)\right) = a + bW_b(z)\ , \qquad [10]$$

for each value of *b* and *z*, thus providing $50 \cdot 5 = 250$ different estimates of the regression coefficients, henceforth denoted by $\theta_b(z)$. For each *z*, the average of the regression coefficients over *b* are calculated as

$$\theta_{\text{Sim}}(z) = \sum_{b=1}^{50} \theta_b(z)/50\,. \qquad [11]$$

Note that $\theta_{\text{Sim}}(z=0) = \theta_w$ and that, for the multiplicative models, $\theta_{\text{Sim}}\left(z = -1/\left(1+\text{Var}_u\right)\right)$ should give a representative estimate of θ_v, which is what we want to find. (For the additive models, $\theta_{\text{Sim}}(z=-1)$ should provide this estimate.) However, from equation [7] it follows that negative values of *z* make $W_b(z)$ complex valued. Since logistic regression analysis can not treat complex valued variables, the SIMEX method instead proceeds by interpolating $\theta_{\text{Sim}}(z)$ over *z* from $z = 0$ to $z = 2.0$ using linear and quadratic regression. The best model (in terms of adjusted R^2) is then used to extrapolate $\theta_{\text{Sim}}(z)$ back to $z = -1/\left(1+\text{Var}_u\right)$ for the multiplicative models and to $z = -1$ for the additive models. This extrapolated value provides the final SIMEX estimate of θ_v.

The SIMEX method was conducted once for each error model. This means that the SIMEX estimate for multiplicative error model 1 provided an estimation of the true fatality risk curve if the estimations errors in GIDAS were correctly captured by that error model. The SIMEX estimates of the other error models should be interpreted analogously.

4. RESULTS

Before proceeding with the results, we recall that the variable *z*, introduced in equation [7], quantified the amount of random error added to the estimated impact speeds, *w*, in the GIDAS database. At *z*=0, no error was added, and the remeasured impact speeds, $W_b(z)$, equaled the estimated impact speeds. When *z* increased, the amount of random error in the remeasured impact speeds increased. Choosing *z* negative could in a statistical sense be interpreted as removing random error from the estimated impact speeds, *w*. However, from equation [7], it followed that negative values of *z* implied complex values for the remeasured impact speeds, which is a remarkable feature. Nevertheless, at a certain negative value of *z*, the SIMEX method can be shown to provide an estimate of the regression coefficients *a* and *b* of equation [1] (or, equivalently, a_v and b_v of equation [5]). Hence, the SIMEX method gives an estimate of what the parameters *a* and *b* would have been if there was no estimation error in the GIDAS database.

4.1 The multiplicative error models

Figure 1 shows how the regression coefficients *a* and *b* in equation [10] varied with *z* for the three multiplicative error models. The best fit regression functions are also displayed in Figure 1. By extrapolating these regression functions back to $z = -1/(1+\mathrm{Var}_u)$, the final SIMEX estimates for a_v and b_v of equation [5] were found (see Figure 1). The corresponding risk functions are provided in Figure 2 together with the results of Ref. [1] (including its 95% confidence interval). It is clear from Figure 1 that a multiplicative estimation error which is normally distributed around the true impact speed implies too high values of the intercept coefficient *a* and too low values of the slope coefficient *b*. This means that such a random estimation error shifts the risk curve towards lower impact speeds and flattens it. From Figure 2, it can be seen that the net effect is to provide slightly too low fatality risks at higher impact speeds.

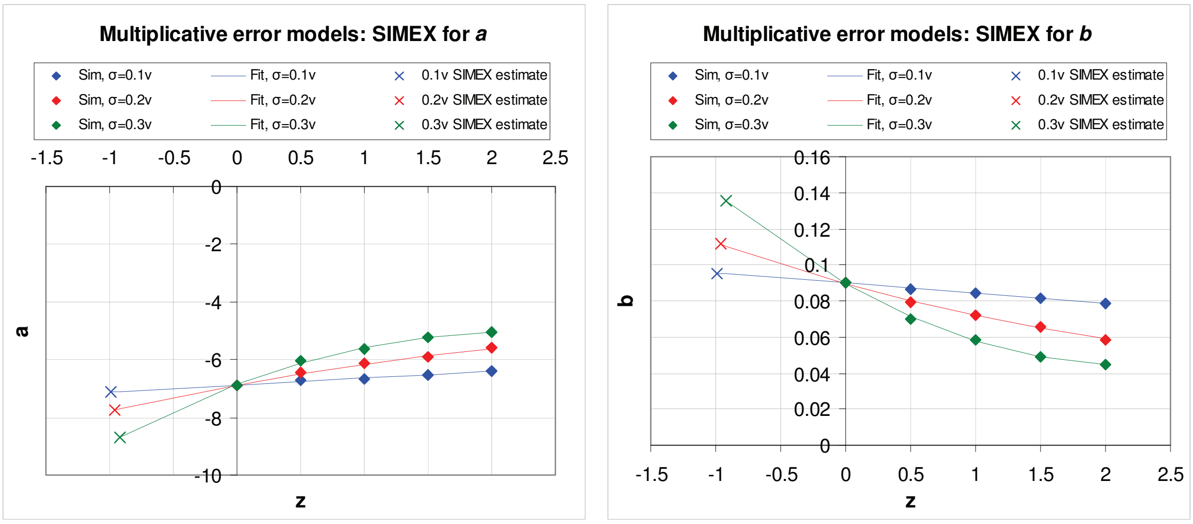

Figure 1. Results from SIMEX analysis for regression parameters *a* (intercept) and *b* (slope) with multiplicative error models.

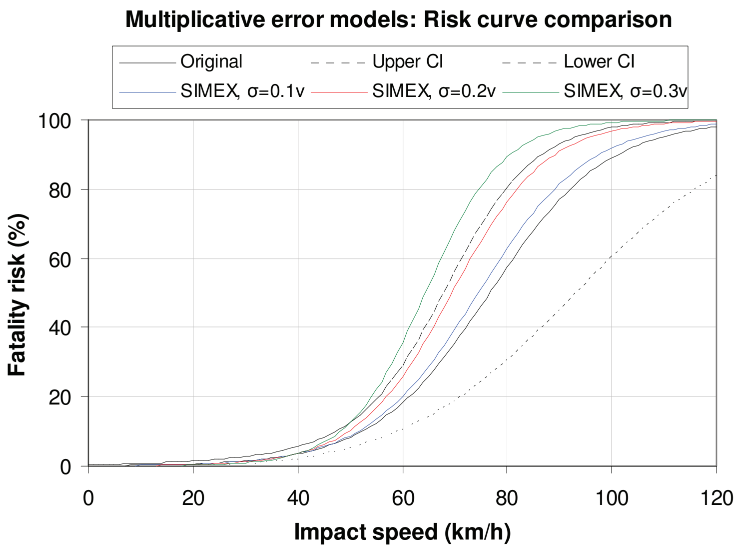

Figure 2. The blue, red, and green curves show the SIMEX estimations of the true fatality risk adjusting for the multiplicative error models described in Section 3.4 (see Table 1). The solid black curve shows the results of Ref. [1] that used the original GIDAS data (dashed curves give 95% confidence intervals).

4.2 The additive error models

Figure 3 shows how the regression coefficients a and b in equation [10] varied with z for the three additive error models. The best fit regression functions are also displayed in Figure 3. By extrapolating these regression functions back to $z = -1$, the final SIMEX estimates for a_v and b_v were found (see Figure 3). The corresponding risk functions are provided in Figure 4 together with the results of Ref. [1] (including its 95% confidence interval). It is clear from Figure 3 that an additive estimation error which is normally distributed around the true impact speed implies too high values of the intercept coefficient a and too low values of the slope coefficient b. This means that such a random estimation error shifts the risk curve towards lower impact speeds and flattens it. From Figure 4, it can be seen that the net effect is to provide slightly too low fatality risks at higher impact speeds. Qualitatively, this is exactly the same effect as that of multiplicative random errors (see Section 4.1).

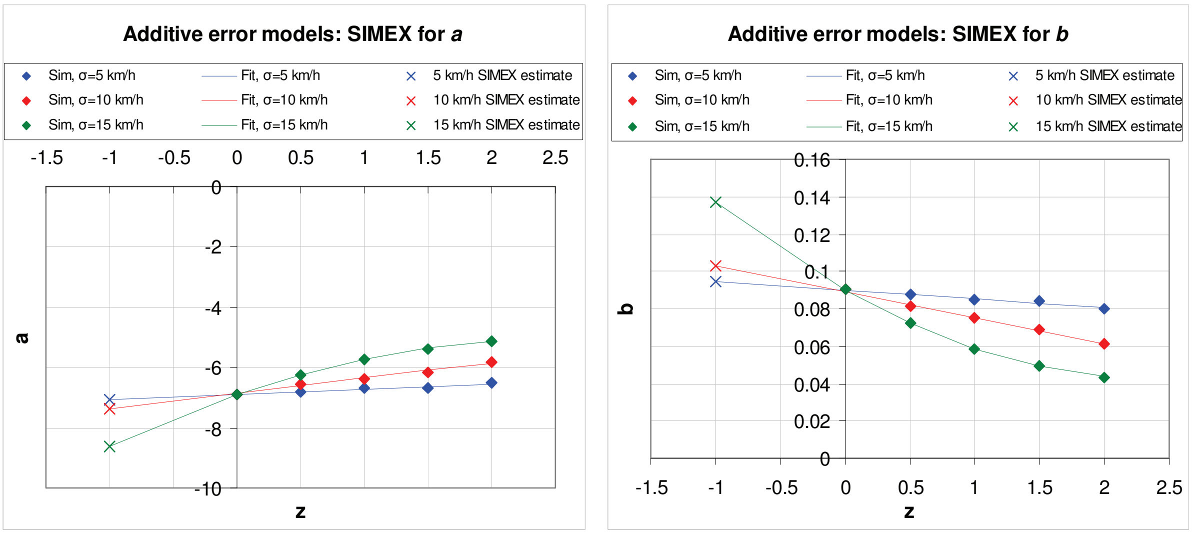

Figure 3. Results from SIMEX analysis for regression parameters a (intercept) and b (slope) with additive error models.

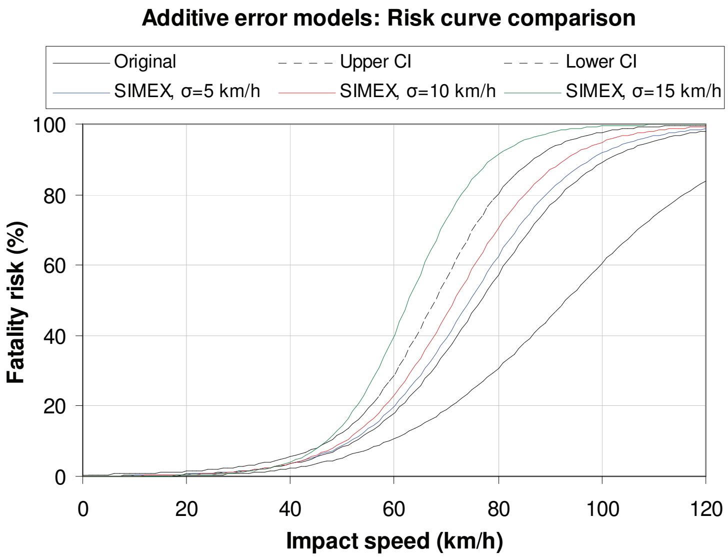

Figure 4. The blue, red, and green curves show the SIMEX estimations of the true fatality risk adjusting for the additive error models described in Section 3.4 (see Table 1). The solid black curve shows the results of Ref. [1] that used the original GIDAS data (dashed curves give 95% confidence intervals).

5. DISCUSSION

In this study, different models for the estimation error of impact speed in the GIDAS database were investigated. Three models assumed a multiplicative error, which means that the standard deviation of the estimation error increased linearly with the true impact speed. The other three models assumed an additive error, which means that the standard deviation of the estimation error was constant for all impact speeds. For all models, the error was assumed to be normally distributed around the true impact speed. The multiplicative models were based on an analysis of the GIDAS reconstruction methodology, see Sections 3.2 and 3.3. In error model 1, the standard deviation, σ, of the random error was 10% of the true impact speed ($\sigma = 0.1v$). Furthermore, $\sigma = 0.2v$ in error model 2 and $\sigma = 0.3v$ in error model 3. Note that 67% of the observations drawn from a normal distribution fall within plus minus one standard deviation (i.e., $v \pm \sigma$ in this study) and 95% of the observations fall within $v \pm 2\sigma$. Hence, taking error model 1 as an example, this means that at a true impact speed of 50 km/h, the estimated impact speed in GIDAS would have 67% probability to fall within 45–55 km/h and 95% probability to fall within 40–60 km/h. Our investigations in Sections 3.3 and 3.4 suggested that different reconstruction methods lead to different estimation errors, but that on average the estimation error in GIDAS was best captured by either multiplicative model 1 or 2. Nevertheless, the tolerance levels given in the GIDAS database suggested that an additive model was more appropriate at impact speeds exceeding 15 km/h, see Section 3.4. This is in line with the theoretical discussion in Section 3.2, that with higher impact speeds more reliable evidence can be collected from the accident site, which leads to an improved quality of the reconstruction. Therefore, three additive models were studied as well, in which the standard deviation of the estimation error was 5 km/h, 10 km/h, and 15 km/h, respectively.

Using the SIMEX method, the pedestrian fatality risk as a function of car impact speed (which was derived in Ref. [1]) was adjusted to take the modelled estimation error into account. It can be seen from Figures 2 and 4 that the multiplicative model with $\sigma = 0.1v$ gave similar results as the additive model with $\sigma = 5$ km/h. The multiplicative model with $\sigma = 0.2v$ gave results comparable to the additive model with $\sigma = 10$ km/h. Finally, the multiplicative model with $\sigma = 0.3v$ gave similar results as the additive model with $\sigma = 15$ km/h. This shows that the qualitative choice of either a multiplicative or additive error model did not have a decisive effect on the results. Let us therefore proceed by discussing, e.g., the multiplicative models: Figure 2 shows that if multiplicative error model 1 ($\sigma = 0.1v$) is correct, the true risk curve is likely to lie close to the curve of Ref. [1]. If error model 2 ($\sigma = 0.2v$) is correct, the true risk curve should still lie within the 95% confidence interval of Ref. [1]. For all error models, the true risk curve is likely to lie close to the curve of Ref. [1] at impact speeds up to 50 km/h. However, at higher impact speeds an adjustment towards higher fatality risks may be motivated. The exact amount of adjustment depends on the size and nature of the estimation error in the data.

Based on these findings, the risk curve of Ref. [1] should provide an accurate description of pedestrian fatality risks up to 50 km/h, but the true risk curve may be slightly steeper at higher impact speeds. These results emphasize the importance of high quality accident reconstructions in order to keep the estimation error of impact speed under control.

Injury risk curves typically follow a logistic distribution function also for other injury levels and for other road traffic casualties than pedestrians. Furthermore, reconstruction methods for vehicle-to-vehicle collisions are based on the same physical principles as car-to-pedestrian crashes. Thus, the qualitative findings of this study can probably be generalised to injury risk curves for other road traffic casualties.

6. LIMITATIONS

The SIMEX method is a tool that can be used to study the effect of error in input data for logistic regression models. It is important to note that it does not provide the absolute truth. Furthermore, a key input to the SIMEX analysis was a statistical description of the estimation error present in the input data, i.e., the GIDAS database in this study. As pointed out in Section 3.3, different reconstruction methods lead to different estimation errors. Hence, using one and the same error model to capture the average estimation error may be a too simplified approach. And even if such an approach is justified, it is quite likely that a multiplicative error model underestimates the error at very low impact speeds and overestimates it at very high impact speeds, and vice versa for an additive model. Furthermore, our assumptions regarding the amount of error involved in the estimations of car brake distance and deceleration, and pedestrian throw distance and deceleration were mainly based on our own experience and common sense instead of observational studies of the GIDAS reconstructions. Finally, our way to apply the SIMEX method for a multiplicative error model is new and has not been critically reviewed.

7. CONCLUSIONS

Car impact speeds estimated from accident reconstructions comprise a certain amount of error. This estimation error has two components: a systematic error and a random error. In this study, the random error was described using both multiplicative and additive error models with different assumptions for the magnitude. Qualitatively, it was found that fatality risk curves for pedestrians are flattened by random estimation error. In particular, the risk becomes too low at higher impact speeds. The flattening effect increased with the size of the error. These findings may be generalised to injury risk curves and to other road traffic casualties as well. It is therefore important that accident investigations and reconstructions are of high quality to assure that systematic errors are minimised and that the random errors are under control. For car-to-pedestrian crashes, the GIDAS database is likely to have only a minor systematic error, while the amount of random error should have a standard deviation less than 15% of the true impact speed. Under these assumptions, it was found that the fatality risk curve derived from the GIDAS database in Ref. [1] is reliable at impact speeds below 50 km/h. At higher impact speeds, a slight modification of the risk curve towards higher fatality risks may be motivated.

ACKNOWLEDGEMENTS

The authors are grateful to Helena Stigson at Folksam Research for introducing us to the SIMEX method and for feedback on the manuscript. Anders Kullgren at Folksam Research and Claes Tingwall at the Swedish Road Administration are acknowledged for discussions on the choice of error models. We also thank Torbjörn Andersson for valuable comments on the manuscript.

REFERENCES

[1] E. Rosén, U. Sander, Pedestrian fatality risk as a function of car impact speed. Accid. Anal. Prev. 41, 536–542, 2009.

[2] E. Rosén, H. Stigson, U. Sander, Literature review of pedestrian fatality risk as a function of car impact speed. Accid. Anal. Prev., doi:10.1016/j.aap.2010.04.003, In press.

[3] R.W.G. Anderson, A.J. McLean, M.J.B. Farmer, B.H. Lee, C.G. Brooks, Vehicle travel speeds and the incidence of fatal pedestrian crashes. In: Proceedings of the 1995 International IRCOBI Conference on the Biomechanics of Impact, Brunnen, Switzerland, pp. 107–117, 1995.

[4] Anderson, R.W.G. Anderson, A.J. McLean, M.J.B. Farmer, B.H. Lee, C.G. Brooks, Vehicle travel speeds and the incidence of fatal pedestrian crashes. Accid Anal Prev. 29, 667–674, 1997.

[5] E. Pasanen, Driving speeds and pedestrian safety; a mathematical model. Helsinki University of Technology, Transport Engineering, Publication 77, 1992.

[6] G.A. Davis, Relating severity of pedestrian injury to impact speed in vehicle pedestrian crashes. Transport Res Rec. Record No. 1773, 108–113, 2001.

[7] L. Hannawald, F. Kauer, Equal effectiveness study on pedestrian protection. Technische Universität Dresden, 2004.

[8] J. Carroll, D. Ruppert, L. Stefanski, C. Crainiceanu, Measurement Error in Non-Linear Models: A Modern Perspective. Chapman & Hall/CRC, 2006.

[9] M. Pfeiffer, J. Schmidt, J., Statistical and methodological foundations of the GIDAS accident survey system. In: Proceedings of the 2nd International Conference on ESAR, Hanover, Germany, pp. 81–87, 2006.

[10] A.J. Dobson, An Introduction to Generalized Linear Models. Chapman & Hall/CRC, 2002.

[11] P. McCullagh, J.A. Nelder, Generalized Linear Models. Chapman & Hall, London, 1989.

[12] W. Hugemann (ed.), Fachbuch Unfallrekonstruktion. Autorenteam GbR, Münster, 2007.

[13] H. Burg, H. Rau, Handbuch der Verkehrsunfallrekonstruktion, Verlag Information Ambs GmbH, 1981.

[14] A. Kühnel, Der Fahrzeug-Fußgänger-Unfall und seine Rekonstruktion. Dissertation, Technische Universität Berlin, 1980.

[15] P. Niederer, M. Schlumpf, Bewegungsmuster angefahrener Fussgängersurrogate, Verkehrsunfall und Fahrzeugtechnik, Heft 2, 1986

[16] D. Otte, Bedeutung und Aktualität von Wurfweiten, Kratzspuren und Endlagen für die Unfallrekonstruktion, Verkehrsunfall und Fahrzeugtechnik, Heft 11, 294–300, 1989.

[17] G. Stcherbatcheff, C. Tarriere, P. Duclos, A. Fayon, C. Got, A. Patel, Simulation of collisions between pedestrians and vehicles using adult and child dummies. In: Proceedings of the 19th Stapp Car Crash Conference. 1 Vol. SAE, Edit., 753–786, 1975.

[18] T. Ziegenhain, Rechnergestützte Untersuchung mittels Mehrkörpermodell in PC-Crash 8.1 zum Zusammenhang zwischen Aufwurf- und Wurfweite eines Fußgängers bei zentralem, gebremsten / ungebremsten Pkw-Frontanstoß unter Variation von Anstoßhöhe, Karosserieform des stoßenden Fahrzeuges, Körpergröße, Kontaktreibung und queraxialer Gehgeschwindigkeit des Fußgängers, in Abhängigkeit von der Kollisionsgeschwindigkeit, Master's Thesis, Technische Universität Graz, 2008.

[19] J.R. Funk, J.M. Cormier, H.C. Gabler, Effect of Delta-V errors in NASS on frontal crash risk calculations. In: Proceedings of the 52nd Annals of Advances in Automotive Medicine, San Diego, CA, 2008.

[20] A.J. McLean, R.W.G. Anderson, M.J.B. Farmer, B.H. Lee, C.G. Brooks, Vehicle travel speeds and the incidence of fatal pedestrian collisions: Volume I. The University of Adelaide, Report No. CR146, 1994.

[21] S.J. Ashton, A preliminary assessment of the potential for pedestrian injury reduction through vehicle design. SAE, Technical Paper 801315, 1980.