

Official reporting and newspaper coverage of road crashes: A case study

Peer-reviewed author version

DANIELS, Stijn; BRIJS, Tom & KEUNEN, Dries (2010) Official reporting and newspaper coverage of road crashes: A case study. In: SAFETY SCIENCE, 48 (10). p. 1469-1476.

DOI: 10.1016/j.ssci.2010.07.007

Handle: <http://hdl.handle.net/1942/11300>

# **OFFICIAL REPORTING AND NEWSPAPER COVERAGE OF ROAD CRASHES: A CASE STUDY<sup>1</sup>**

Stijn Daniels\*, Tom Brijs, Dries Keunen

Hasselt University - Transportation Research Institute

Wetenschapspark 5 bus 6, 3590 Diepenbeek - Belgium

e-mail [stijn.daniels@uhasselt.be](mailto:stijn.daniels@uhasselt.be); [tom.brijs@uhasselt.be](mailto:tom.brijs@uhasselt.be)

tel +32 11 269111 fax +32 11 269199

\*Corresponding author

Keywords: crash, reporting, rate, newspaper, coverage

<sup>1</sup> Paper accepted for publication in Safety Science

## **ABSTRACT**

This paper describes a case study in Flanders-Belgium on the reporting rate of road crashes. Crash data from three sources were compared: official crash data, data retrieved from an insurance company and newspaper articles. A sample of 140 injury crashes with motorcyclists from an insurance company was used as the reference category. The purpose was to explore factors that contribute to the likelihood of crashes (not) to be reported in official statistics and newspapers. Logistic regression analyses and chi-square tests were used to reveal differences in reporting rate according to some variables. About 80% of the crashes with severe injuries were reported in the official statistics whereas the reporting rate for crashes with slightly injured was about 55%. Newspapers covered about 50% of crashes with severe injuries. The reporting rate in both official statistics and newspapers increased with the severity of the crash.

# 1. INTRODUCTION

## 1.1. Problem

Approximately three persons a day are killed in road crashes in Belgium. 20 persons get severely injured and 160 people get slightly injured (Statistics Belgium, 2007). Crash statistics are one of the most important data sources for road safety policy. Consequently, their correctness might be influential to policy evaluation and to the perception of the importance of the road safety problem (Evans, 2004). Road infrastructure investment programs use to be based – at least partly - on the official numbers of crashes at particular locations. Obviously, such a policy assumes accurate crash data.

However, the collection of road crash data might be biased or incomplete for a plenty of reasons. One important reason is the underreporting of crashes, which means that crashes that should be included in official databases are not. Underreporting can be related to a number of reasons (Elvik & Vaa, 2004; Hauer, 1997):

- Police intervention is not asked and even not legally required for each road crash, e.g. not for crashes with property damage only nor for crashes with only pedestrians. Without police intervention, no official reporting of a crash can happen.
- Some information may get lost in the reporting process or might be incorrectly provided for various reasons such as misclassification of injury severity, lack of information on seat belt wearing or alcohol use, missing location data and encoding errors.

A correct reporting of crashes is believed to be of utmost importance since many policy measures and evaluation studies largely rely on the quality of the underlying crash counts. Not only the simple reporting rate of the entire mass of crashes is important, but also the question whether this rate is structurally dependent of variables like the involved types of road users (pedestrians, car drivers, bicyclists...), time of the day, day of the week, injury severity, age and gender of the involved. Under-reporting is defined as the share of crashes that ought to, but are not reported in the total number of reportable crashes. For example, a higher rate of under-reporting for crashes with some user categories might lead to a systematic underestimation of the magnitude of the safety problem for those categories.

The purpose of the present paper is to examine the reporting rate of road crashes in Belgium by means of a case study in which crash data from three sources were compared: official crash data, data retrieved from an insurance company and newspaper articles.

## 1.2. Crash Reporting Procedure

Injury crashes in Belgium are legally required to be reported to the police. Reporting is optional in case of property damage only. Information on injury crashes is registered in a crash reporting form that is an obligatory attachment to the official minutes of the crash. After completion, the crash reporting form is transferred by the police to the Federal Public Service Economy, that takes care of the aggregation and dissemination of the results. In case of fatalities or injuries that result in a fatality within a period of 30 days after the crash, a supplementary data stream is created since the Federal Public Service Justice delivers routinely mortality data to the Federal Public Service Economy. Therefore, the fatality data are double-checked. No information on property damage only-crashes is collected at the national level.

In the official data, a distinction is present between three types of injuries: killed (on the spot or within 30 days after the crash), seriously injured (at least 24 hours in a hospital) and slightly injured (injured and not belonging to one of the two previous categories).

### 1.3. Media Reporting

Media are playing an important role in the creation of health awareness, e.g. road safety attitudes in a population (Connor & Wesolowski, 2004). One aspect of this health awareness relates to the perceived reasons of mortality. Frost et al. (1997) suggest that mass media are creating a biased perception of different causes of death. The number of reported crashes and the way of reporting are likely to be influential to the perception of the risk of certain behaviors. Whenever this risk does not correspond with the real risk, biases in the perception of the risk might occur. Traffic crashes just like other events are more likely to get reported by news media when they are newer, more rarely occurring or more dramatic (Adams, 1992). The seriousness of the problem is not necessarily a determining factor. Consequently, it is possible that people have a wrong or biased perception of the size of the injury risk in traffic.

In the present study, the official reporting rate of a sample of road crashes is checked. Moreover, the coverage of the sample crashes in newspapers is examined and some checks were done on the correctness of the presented information.

## 2. DATA

Data from three different sources were collected: an insurance company, the official database and an electronic media archive.

### 2.1. Insurance Data

A sample of insurance data on 140 injury collisions with at least one motorcyclist was provided by an insurance company and served as a reference category. This sample was previously used for an in-depth analysis of motorcycle crashes (Van Hout, 2007). The information in the dataset was obtained from several documents: the official minutes of the crash like recorded by the police, the collision declaration filled out by the involved parties and medical information such as a doctor's certificate containing information about the nature and severity of the injuries. In a number of cases, supplementary information was present from an insurance inspector's field report. Furthermore, in a limited number of cases, mainly in case of fatalities or critically severe injuries, a detailed report of a forensic expert was available.

The examined crash files were those where the opponent (regarded from the viewpoint of the insurance company's customer) was a motorcyclist. A consequence of this approach was that all included crashes were crashes with at least two road users. Single-vehicle-crashes are therefore not present in this database.

The data covered the period between 1991 and 2005. All selected crashes occurred in the Flanders region in Belgium.

### 2.2. Official Data

Official crash data were retrieved from the detailed database of road crashes in Belgium (Statistics Belgium, 2007) and were available for the period 1991-2005.

### 2.3. Newspaper Articles

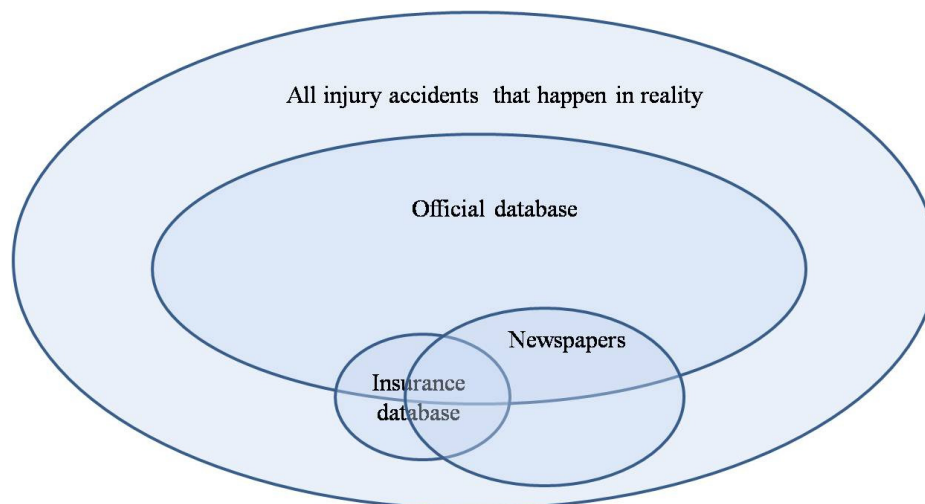
Information on the coverage of road crashes in written media (newspapers) was retrieved from Mediargus ([www.mediargus.be](http://www.mediargus.be)), a professional information provider, which includes the Dutch speaking Belgian newspapers. Mediargus covers articles that were published since 1988. An implicit assumption was that the Mediargus database covers the content of the included newspapers completely and correctly with respect to this issue.

## 3. METHOD

Figure 1 reflects the conceptual relation between the three different data sources. The insurance data act as reference category. The insurance data were for the purposes of the previous in-depth analysis (Van Hout, 2007) thoroughly checked and cleaned. Consequently, the reliability of the resulting information concerning date, location and severity of the injuries in the crash was considered high.

In a first step, it was checked whether the crashes that occurred in the insurance database were also present in the official crash data. The data were matched manually, based on the available information on date and location of the crashes in both datasets. In case of uncertainties, also information about age and gender of the involved was used. The resulting information was stored in a spreadsheet.

Subsequently, the same principles were applied to perform the media search. The following information was retrieved from the newspaper articles: reported crash cause(s), severity of the injuries, number of involved, age and gender, vehicle/road user type.



**FIGURE 1 Conceptual relation between different data sources.**

The official reporting rate and the media coverage rate are expressed as follows:

$$\text{Official reporting rate (R}_{\text{off}}\%) = (\text{ACC}_{\text{off}}/\text{ACC}_{\text{ins}}) \times 100 \quad (1)$$

$$\text{Media coverage rate (R}_{\text{med}}\%) = (\text{ACC}_{\text{med}}/\text{ACC}_{\text{ins}}) \times 100 \quad (2)$$

with  $\text{ACC}_{\text{off}}$  = number of crashes in the insurance database that were traced in the official database

$\text{ACC}_{\text{ins}}$  = number of crashes in the insurance database

$\text{ACC}_{\text{med}}$  = number of crashes in the insurance database that were traced in the media database

Subsequently chi-square tests and logistic regression analyses were carried out to reveal influential variables on the likelihood of crashes to be reported. Included variables were gender, age, day and time of the crash, injury severity and type of road user.

## 4. RESULTS AND DISCUSSION

The results of the analyses are provided hereunder, beginning with the crash reporting in the official statistics and subsequently for the media coverage. Since these were two relatively distinct parts, we considered it useful to present the results together with the discussion elements for each of them.

### 4.1 Crash Reporting in the official statistics

140 injury crashes from the insurance database were matched with the official crash data. 86 of the 140 crashes could be traced in the official database, which equals a reporting rate of 61%. Subsequently, some more details were investigated by splitting up information according to injury severity, age and gender of the involved, time and date of the crash, and road user category. Bivariate analyses were performed by chi-square tests that revealed some possible structural relationships between certain variables and the likelihood of reporting. Furthermore, a logistic regression model was fitted for the likelihood of a given crash to be reported as a function of some explanatory variables. Table 1 lists the available variables.

**TABLE 1 Explanatory Variables**

Variable	Frequencies in the insurance dataset (N=140)
Severity	Fatal: 5; Severe injuries: 40; slight injuries: 95
Gender motorcyclist	Male: 126; Female: 13; Unknown: 1
Gender opponent	Male: 107; Female: 33
Age motorcyclist	16-21: 38; 22-30: 27; 31-40: 27; 41-50: 29; 51-60: 7; 60+: 6; Unknown: 6
Age opponent	16-21: 10; 22-30: 26; 31-40: 31; 41-50: 31; 51-60: 13; 60+: 25; Unknown: 4
Day of the week	Mon: 17; Thu: 18; Wed: 20; Thu: 20; Fri: 24; Sat: 20; Sunday: 21
Week/weekend	Week: 99; Weekend (sum of Saturday and Sunday): 41
Time	00-02: 3; 02-04: 0; 04-06: 0; 06-08: 19; 08-10: 9; 10-12: 12; 12-14: 16; 14-16: 19; 16-18: 36; 18-20: 15; 20-22: 6; 22-24: 3; Unknown: 2
Daytime/nighttime	Day (06-22):134 ; Night (22-06): 6
Month of the year	Jan: 9; Feb: 5; Mar: 13; Apr: 13; May: 11; Jun: 15; Jul: 12; Aug: 11; Sep: 18; Oct: 11; Nov: 12; Dec: 10
Road user type (opponent)	Car: 123; Bicycle/moped: 4; Motorcycle: 6; Truck: 4; Unknown: 3
Year of the crash	1998: 1; 1999: 1; 2000: 4; 2001: 2; 2002: 9; 2003: 22; 2004: 44; 2005: 57

A forward stepwise logistic regression procedure was adopted, meaning that a stepwise model was created beginning from zero. Subsequently all individual variables were checked and each time the most significant variable was allowed to enter. Alternative models were compared, based on their reasonableness and their goodness of fit, using the Akaike Information Criterion (AIC) and the Bayesian information criterion (BIC). Because the penalty of adding regressors is higher for the BIC, this criterion tends to favour more parsimonious models than AIC (Verbeek, 2004). It was chosen to minimize the BIC-criterion since the sample size (N=140) was limited and quite some candidate regressors were present, a case in which the AIC might behave poorly. The resulting model is shown in table 2.

The third column of table 2 shows the odds-ratios for the different variables. An odds-ratio provides the most convenient way to interpret the model results. As all the variables were expressed as indicator variables, the interpretation of the odds-ratio is that when the concerned variable is present in the crash (e.g. Year = 2000), the odds-ratio reflects the relative probability of the crash to be reported for the considered case, compared with the case in which the condition is not true. Crashes on Saturdays appear to be four times more likely to be officially reported (odds-ratio of 4.04) than crashes on other days. Table 2 shows that also crashes with serious injuries (fatalities or severe injuries) and crashes where the motorcyclist collides with a car are more likely to be reported than crashes with only slight injuries or collisions with other road users. Furthermore, crashes in 2000 were substantially less likely to be reported than crashes in other years (OR=0.07).

The Hosmer and Lemeshow goodness-of-fit test (Hosmer & Lemeshow, 1989) indicates no significant differences ( $\chi^2 = 0.59$ , df 5, p 0.99) between the fitted model and the observed data, meaning that the model matches the data correctly.



**TABLE 2 Logistic Regression Results for Official Reporting Rate**

Variable	Parameter estimate (p-value)	Odds Ratio	OR 95% Wald confidence limits	
Intercept	-1.12 (0.06)			
Day (Saturday vs. other)	1.40 (0.04)	4.04	1.05	15.58
Injuries (Fatal or Severe vs. Slight)	1.12 (0.01)	3.06	1.29	7.27
Year (2000 vs. other)	-2.68 (0.04)	0.07	0.01	0.89
User type opponent (Car vs. other)	1.33 (0.03)	3.79	1.13	12.70

Hosmer and Lemeshow test  $\chi^2 = 0.28$  (df = 3, p = 0.96), BIC 190.75, generalized  $R^2 = 0.14$

A series of bivariate chi-square tests for the relationship between each individual variable and the reporting rate were done as well. Only one variable had an individual significant effect on the reporting rate: the injury severity ( $\chi^2 = 8.26$ , df 2, p = 0.02).

#### 4.1.1 Injury Severity

Only two of the five fatalities in the sample appeared to be officially reported, which is surprisingly low (Table 3). A previous Belgian study, based on a comparison of 120 judicial records with the official crash data, showed a reporting rate for fatalities of 92% (Mens en ruimte, 1997). Elvik & Mysen (1999) made a meta-analysis of 49 studies on crash reporting in 13 different countries and found an average reporting rate for fatalities of 95%.

Some chance elements, related to the small size of the sample, might be assumed to explain the poor registration of the fatal crashes. However, the statistical likelihood of at least 3 misses in a small sample of 5 observations, given an assumed population proportion of 0.92 successes is less than 0.5% and is therefore extremely low.

**TABLE 3 Reporting Rate and Media Coverage Rate**

Injury severity <sup>1</sup>	N	Official reporting		Media coverage	
		n	%	n	%
Fatalities	5	2	40%	3	60%
Crashes with seriously injured	40	32	80%	21	52%
Crashes with slightly injured	95	52	55%	17	18%

<sup>1</sup>severity of the most serious injury in the crash

The observed reporting rate of crashes with severe injuries in the present study was 80%. Table 4 shows results for under-reporting of severely injured in previously executed studies on crash reporting in Belgium. These studies were either university-based or carried out by a government agency. Some difficulties in comparing and interpreting their results could arise since the used methods and data sources were different. Aelvoet et al. (2002) compared data in the 'Minimal Clinical Data' database that were classified as road crashes with the official crash data on severely injured that were provided by the National Statistical Institution based on police data. They found a reporting rate of 38.7-42% for this crash category. De Mol (2002) found a reporting rate of 63% for severe crashes based on the information in questionnaires completed by a convenience student sample of 2900 subjects. Lammar & Hens (2006) carried out a case study of crash reporting based on clinical data of two hospitals. They reported a registration rate of 43-60% of the crashes.

The present study finds a higher reporting rate, which could be related to several reasons:

- The nature of the included crashes. The present sample includes only multiple vehicle collisions, which is likely to affect the likelihood of being reported positively (see 4.1.4). Furthermore, the present sample includes only crashes with motorcyclists, which also might affect the reporting rate. However, the reporting rate for motorcycle crashes appeared sometimes to be lower than for other user types (see the discussion in 4.1.5).
- Differences in the used data sources. The present study used insurance data, which was not the case elsewhere in the reported studies. Some other studies used hospital data, one used local police data, one used self-reported data by means of a survey among students. Biases in each of these data sources might exist for different reasons (Shinar, 2007).
- Differences over time: it might be possible that the data collection system improved during the past decade, e.g. related to computerization of procedures.

**TABLE 4 Study Results on Reporting Rate of Severely Injured in Belgium**

Author(s)	Data source	Year of crash data	N	Reporting rate
Aelvoet et al. (2002)	Hospital data	1996	6 658	42%
		1997	7 492	41.7%
		1998	8 188	38.7%
De Mol (2002)	Self-reported data	2002	2 900	63%
Lammar & Hens (2006)	Hospital data	2000 - 2002	498	53% - 60%
		2000 - 2002	518	43% - 50%

The reporting rate for crash with slight injuries in the present study was 55%.

#### **4.1.2 Age**

The age of the involved did not seem to be influential. A Dutch study showed some differences in reporting according to age, but mainly crashes with children in the age category 0-14 were reported less (Mulder et al., 1995). However, this last result could partly relate to the fact that children are more involved in pedestrian and bicycle crashes that are less likely to be reported (James, 1991). No other clear indications on possible age effects are available from the examined literature.

#### **4.1.3 Gender**

Neither the gender of the motorcyclist nor the gender of the opponent appeared in the final model. A forced inclusion in the model did not seriously affect the estimates for the other parameters and showed a tendency towards a lower reporting in case of a male motorcyclist (OR= 0.36 with  $p=0.16$ ) and in case of a male opponent (OR= 0.54,  $p=0.20$ ). Moreover, an interaction effect was included to model the case that both parties were male, but this added nothing to the model. Amoros et al. (2006) found a slightly lower reporting rate for crashes with females. Alsop & Langley (2001) found no difference according to the gender of the involved.

#### **4.1.4 Single/multiple Vehicle Crashes**

Some studies distinguished also between single and multiple vehicle crashes (Alsop & Langley, 2001; Amoros et al., 2006; Elvik & Mysen, 1999). They reported consistently a lower reporting rate for single vehicle crashes than for multiple vehicle crashes. In our sample, only multiple vehicle crashes were included. This might explain partially why the reporting rate in our study appeared to be relatively high.

#### **4.1.5 Road User Type**

Alsop & Langley (2001) examined the influence of the road user type on the reporting rate of the concerning crash. Crashes with motorcyclists or moped riders showed a lower reporting rate than crashes with other vehicle drivers (60% vs. 70%). James (2001) reported a generally higher level of reporting for car/van crashes compared with motorcycle crashes. Elvik & Mysen (1999) found similar results in their meta-analysis: the highest reporting levels for car occupants, generally slightly lower for pedestrians, still lower for motorcyclists and the lowest for bicyclists. However, the analysis by Amoros et al. (2006) did not confirm the lower reporting rate for crashes with motorcyclists in multiple-vehicle collisions, the case that is applicable to the data in the present study. They found a reporting rate that was very similar to that of car occupants (in fact somewhat higher, but not significantly).

In the present study, the reporting rate appeared to be higher when the motorcyclist collided with a car than when the opposite party was no person car.

#### **4.1.6 Time Aspects**

The examined data showed some differences in the reporting rate, according to the year and the date of the crash. Crashes in 2000 were less likely to be reported, but inspection of the frequency data revealed no clear pattern during the subsequent years. The reporting of crashes on Saturday appeared to be higher. The hour of the day did not seem to have an impact. Alsop & Langley (2001) found no effect of the month of the year, but they reported a lower reporting rate on Sundays. Also in the present study, the reporting rate was the lowest on Sundays, although it differed not significantly from other days. Limitations in the capacity of police forces on Sundays could explain lower reporting rates, but the more it remains unclear why the reporting rate seems to be particularly higher on Saturdays.

### **4.2 Media coverage**

#### **4.2.1 Reporting rate**

The 140 injury crashes in the insurance database were matched one by one with the data of the online media provider Mediargus. It was checked for each crash whether it was reported in the media database. For each crash in the insurance database, a search in the media database was done with a scope of 1-5 days after the day of occurrence of the crash. Used search criteria in the media database were 'crash', 'motorcycle' and 'motorcyclist'. Subsequently, crashes were selected that occurred in the same municipality as the crash in the insurance database.

After this selection, the remaining articles were checked on their accordance with the crash in the insurance database, based on variables where information was available for from the insurance data like the number and gender of the involved and the types of involved road users.

41 of the 140 crashes were traced back in the Mediargus database, which equals a reporting rate of approximately 29%. It was assumed that Mediargus covers the content of the included newspapers completely. If, for whatever reason, this assumption would not hold, the revealed reporting rate should be considered as a minimum.

Again, a series of bivariate chi-square tests for the relationship between each individual variable and the reporting rate was executed. Two variables had an individual significant effect on the reporting rate: the injury severity ( $\chi^2 = 18.62$ , df 2,  $p < 0.01$ ) and the day of the week ( $\chi^2 = 12.88$ , df 6,  $p = 0.04$ ).

Three out of the five fatalities were reported in the newspapers and 52% of the crashes with severely injured were reported. In case of crashes with slightly injured, the likelihood of being reported was only 18% (Table 3).

Furthermore, a logistic regression model was fitted for the likelihood of a crash to be reported in at least one of the included newspapers as a function of some possible dependent variables. The captured data were the severity of the worst injury in the crash, the gender of the motorcyclist and the other involved party, the age of the involved, information on the time and day of the crash and the road user type of the opposite party (see Table 1). The provided information in the newspaper articles on possible contributing factors to the crashes (such as alcohol use, overtaking or failed to yield) was captured as well, but not integrated in the quantitative analysis since none of these factors turned out to occur sufficiently frequently for quantitative analysis (see Table 7).

A forward stepwise logistic regression model was applied, following the same procedure as mentioned above. The resulting model is shown in table 5.

**TABLE 5 Logistic Regression Results for Media Reporting**

Variable	Parameter estimate (p-value)	Odds Ratio	OR 95% Wald confidence limits	
Intercept	-3.40 (0.01)			
Day of the week (Saturday vs. other)	1.31 (0.02)	3.70	1.19	11.50
Injuries (Fatal or Severe vs. Slight)	1.72 (<0.01)	5.60	2.42	12.99
Daytime vs. Nighttime	1.59 (0.21)	4.90	0.40	59.34
Year (2002 vs. other)	1.73 (0.04)	5.64	1.09	29.10

Hosmer and Lemeshow test  $\chi^2 = 1.33$  (df = 3, p = 0.72), BIC 165.18, generalized  $R^2 = 0.19$

The Hosmer and Lemeshow goodness-of-fit test indicates no significant differences between the fitted model and the observed data, meaning that the model matches the data adequately.

The results of the logistic regression show that the reporting rate in the media is substantially higher in case of serious injuries or fatalities, which was not unexpected. Neither age and gender of the motorcyclist nor age and gender of the involved motorcyclist appeared to be influential.

More important were time aspects like day of the week, daytime versus nighttime and year. Although not significant and therefore only highly indicative, the weaker reporting during nights might be explained by less available resources during these periods. This might be an important element, since it could show a structural bias in information processing on crashes by written mass media.

The presence of a particular year variable (2002) might indicate some fluctuations in mass media attention for the social problem of traffic crashes like was described by Hilgartner & Bosk (1988) in their Public Arenas Model. According to this model, the amount of public attention or media attention for a certain social problem could be structurally related to some factors like the selection criteria that newspapers use to report about certain events or to the dynamics of the competition between different social problems. Applied to the present case, some particular reasons could explain why more attention went perhaps to traffic crashes in 2002. A search in the Mediargus database reveals that 2002 was in Belgium characterized by a lively public debate on the installation of a series of automated speed cameras, the height of fines in traffic and by the political announcement of a speed limit reduction. Therefore, it is plausible that the magnitude of road safety problems was perceived to be higher and more attention in the media was given to road safety items. However, a detailed data

analysis for the effects of the subsequent years did not reveal a clear increasing or decreasing pattern in the reporting rate, which would have supported the thesis of structural evolutions.

Comparable mechanisms could explain the higher reporting rate on Saturdays. There are relatively few news items on Saturdays compared to weekdays, which makes it easier for crashes to ‘compete’ with other events or social phenomena for attention in the mass media. Moreover, newspaper offices might be better prepared to report about road traffic crashes on Saturdays, since weekend crashes tend to be perceived as an important social problem. However, one should interpret these findings with much caution due to the small size of the dataset, which makes the occurrence of some chance effects more likely and might conceal other effects.

#### 4.2.2 Reporting quality

Apart from checking the mere fact of reporting the crash in the media, the correctness of the reporting was examined as well. This was done by comparing a number of variables that were reported in the media with some features of the insurance database. Used variables were the injury severity, the number of involved in the crash, the number of passengers, age and gender of the involved, vehicle type and finally the mentioned cause of the crash. 22 of the 41 newspaper articles contained information about the severity of the crash (Table 6). For each of these variables, it was checked whether the description in the newspaper articles matched the information in the insurance database. However, this happened without further discourse analysis, see some comments below. The presented information was correct in 18 of the 22 cases (82%) whereas four articles contained incorrect information. The number of involved road users was correctly mentioned in each of the 41 articles. The road user category of the opponent was also nearly always mentioned and if so, it was done correctly. Age and gender were often not mentioned, but if they were, the presented information was mostly correct. In 17 of the 26 newspaper articles where age was mentioned, only the age of the motorcyclist was reflected whereas 9 articles mentioned the age of both the motorcyclist and the opponent.

**TABLE 6 Factual correctness of Media Reporting**

	Reported correctly	Not correctly reported	No information
Injury severity	18	4	19
Number of involved	41	0	0
Gender of the involved	18	2	21
Age of the involved	25	1	15
Road user category (opponent)	40	0	1

N= 41 (crashes reported in the newspapers)

Table 7 reflects the mentioned contributing factors in the newspaper articles. It shows that in a vast majority of the cases no factor is indicated.

**TABLE 7 Frequency of Mentioned Crash Factors in Newspaper Articles**

Causes	n	%**
Alcohol	1	2.44%
Evading maneuver	1	2.44%
Failed to yield	2	4.88%
Overtaking maneuver	4	9.76%
Motorcyclist not visible	4	9.76%
No causal factor	30	73.17%
Total	42*	100%

\* Total does not correspond with total of newspaper articles since in some articles more than one (causal) factor was mentioned.

\*\* Percentage of the total number of newspaper articles (N = 41)

These figures correspond also quite well with a previous study about traffic crash reporting on TV channels in Flanders-Belgium (Beullens et al., 2008) where was found that in 67% of the cases no causes for the crash were mentioned.

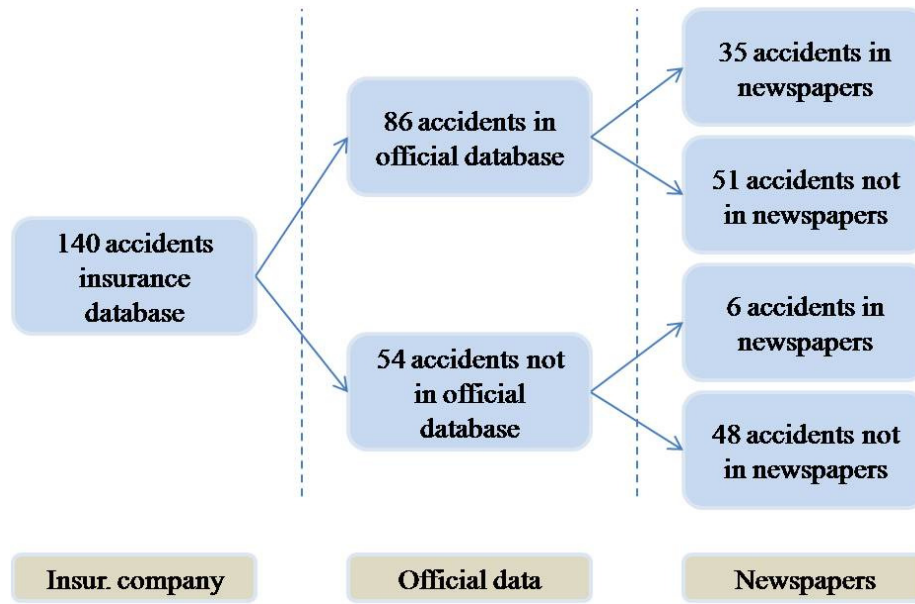
Since the data in the insurance database were enriched with reliable background data of the contributing crash factors, a useful comparison was possible between the insurance data and the reported contributing crash factors in the media database. In 8 of the 11 newspaper articles (73%) where sufficient information was available for, the most important factor was reflected correctly.

These results seem to indicate that the factual information in the newspapers concerning traffic crashes is generally correct. However, the mere fact that some elements were mentioned correctly or even the fact that important crash factors were in most cases correct when they were mentioned, does not guarantee that all relevant elements to the public perception of traffic risk are adequately reflected. In order to acquire deeper insights in the way how mass media are reporting about a certain phenomenon, the technique of framing research (Scheufele & Tewksbury, 2007) is likely to be appropriate. Framing is based on the assumption that *how* a message is characterized in media reports is important in the way in which this message is perceived. The reader should notice that, in the context of the present study, no framing research was elaborated. Our approach was rather the one of the *agenda setting* model (Scheufele & Tewksbury, 2007), meaning that was investigated *whether* some events were reported in the written mass media, without exploring *how* this was reported. Beullens et al. (2008) investigated television news' framing of motor-vehicle crashes. They indicated that media seem to personalize and emotionalize traffic crashes, which might stress some aspects of crashes while neglecting others such as protective measures like seat belt wearing. Another instance of research on road crash framing is found in Connor & Wesolowski (2004) who published a study on the newspaper coverage of traffic crashes in four Midwestern cities in the United States. They made a comparison between newspapers and the official accident reporting system (FARS).

Evidence for a particular over-representation of crashes with young people like suggested in Beullens et al. (2008) and in Connor & Wesolowski (2004) was not found in the present data.

### 4.3 Link between the Three Databases

Figure 2 reflects some links between the three databases. It makes clear that a higher reporting rate in the official database correlates with a higher reporting rate in the newspapers. This is confirmed by an additional logistic regression analysis in which the model in table 5 was extended with a covariate to indicate the fact whether the crash was officially reported or not. This variable showed an individual odds-ratio of 4.14 (relative chance of being reported in the media) with a significance level of 0.01.



**FIGURE 2 Link between insurance data, official data and media reporting.**

## 5. STUDY VALIDITY

Both internal and external validity issues are of importance. The internal validity refers to the degree in which the results reflect the reality for the included 140 crashes in the dataset. Since the data were carefully checked and appropriate statistical analyses were applied, it should warrant that the results are reflecting the reality for the investigated crashes.

As to the external validity, several cautions must be given. Firstly, the crashes refer only to crashes that happened in one region in one country (Flanders-Belgium). Secondly, only multiple vehicle collisions with at least one motorcyclist were considered. Thirdly, the examined sample of 140 crashes is relatively small. Fourthly, we disposed of data from only one insurance company. For those reasons, the present study should be qualified as a case study.

We took the opportunity to use insurance data since this type of data was hitherto rarely used in traffic safety research, although it was believed to be potentially important (Elvik & Mysen, 1999; James, 1991). According to the latter authors, insurance data could mainly serve as a complement to other data sources such as hospital data in order to estimate the level of under-reporting. Elvik & Mysen (1999) list a number of sources of road crash data - police records, hospital records, insurance and other company records and self-reported data - and suggest that the most complete coverage of accidents would be achieved by combining data from all those sources.

Our approach allowed a comparison of three databases, i.e. the insurance data, the official crash data and the newspaper articles, which revealed some interesting structural relationships between those data. This was particularly interesting since those databases are maintained independently from each other, in the sense that their data collection procedures are different and they do not rely on observations by the same people.

## 6. CONCLUSIONS

Crash data from three sources were compared: official crash data, data retrieved from an insurance company and newspaper articles. A sample of 140 injury crashes with motorcyclists from an insurance

company acted as the reference category. The purpose was to explore factors that contribute to the likelihood of crashes (not) to be reported in official statistics and newspapers.

About 80% of the crashes with severe injuries appeared in the official statistics, which is higher than was found in previous studies. The reporting rate for crashes with slightly injured was about 55%. Only 2 of the 5 fatal crashes were traced in the official database, which is a counterintuitive result. The reporting rate varied according to the injury severity (higher for the most serious injuries), day of the week (higher on Saturdays), year (lower in 2000) and opponent vehicle type (higher for cars), but the data did not enable to provide a straightforward explanation for these effects. The reporting rate for crashes with males was somewhat, but not significantly, lower than for crashes with females.

Newspapers covered about 50% of crashes with severe injuries. Like expected, the media coverage rate was substantially higher for the most severe crashes. Neither age and gender of the motorcyclist nor age and gender of the involved opponent appeared to be influential. More important were time aspects like day of the week, daytime versus nighttime and year. Although not significant and therefore only highly indicative, the weaker reporting during nights might be explained by less available resources at night. This might be an important element, since it could show a structural bias in information processing on traffic crashes by written mass media.

No statements can be made about the way media are representing crash information. Newspaper articles rarely specify possible crash causes, but provide information on age, gender, number of involved, injury severity and road user type of the involved. Although the latter information appeared to be generally correct, the study design does not allow assessing how newspapers readers perceived the presented information. Further research on this topic is recommended and should incorporate the framing of road crashes in the written mass media.

The presented results rely on an examination of a limited number of multiple-vehicle crashes with one particular road user group: motorcyclists. Consequently, the results are only indicative. Repeating this research with larger and more randomized datasets could, in addition to information obtained from other data sources, contribute to the knowledge on the level of under-reporting for crashes.



## 7. REFERENCES

- Adams, W.C. (1992). The role of media relations in risk communication. *Public relations Q.*, 37, 28-32.
- Aelvoet, W., Mertens, I., Van Sprundel, M. (2002). Uitwendige oorzaken in de Minimale Klinische Gegevens. Federale Overheidsdienst Volksgezondheid.
- Alsop, J., Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis and Prevention*, 33(3), 353-359.
- Amoros, E., Martin, J., & Laumon, B. (2006). Under-reporting of road crash casualties in France. *Accident Analysis & Prevention*, 38(4), 627-635.
- Beullens K., Roe K., Van den Bulck J. (2008). Television news' coverage of motor-vehicle crashes. *Journal of Safety Research*, 39, 547-553.
- Connor, S.M., Wesolowski, K. (2004). Newspaper framing of fatal motor vehicle crashes in four Midwestern cities in the United States, 1999 - 2000. *Injury Prevention*, 10, 149-153.
- De Mol, J. (2002). Opteren voor OPTIMA. Betrouwbaar registratiesysteem voor verkeersslachtoffers. *Verkeersspecialist*, 87, 5-8.
- Elvik, R., Mysen, A.B. (1999). Incomplete accident reporting. Meta-analysis of studies made in 13 countries. *Transportation Research Record*, 1665, 133-140.
- Elvik, R. & Vaa, T. (2004). The handbook of road safety measures. Elsevier, Oxford, UK.
- Evans, L. (2004). *Traffic Safety*. Bloomfield Hills, Michigan, USA: Science Serving Society.
- Frost, K., Frank, E., Maibach E. (1997). Relative Risk in the News Media: A Quantification of Misrepresentation. *American Journal of Public Health*, 87(5), 842-845.
- Hauer, E. (1997). *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Oxford: Pergamon Press.
- Hosmer W. D., Lemeshow S. (1989). *Applied Logistic Regression*. New York: Wiley.
- James, H.F. (1991). Under-reporting of accidents. *Traffic engineering and control*, 32(12), 573-580, 583.
- Hilgartner, S., & Bosk, C. (1988). The Rise and Fall of Social Problems: A Public Arenas Model. *American Journal of Sociology*, 94(1), 53-78.
- Lammar, P. & Hens, L. (2006). Casestudies onderregistratie van ernstig gewonde verkeersslachtoffers. Officiële ongevalgegevens versus ziekenhuisgegevens. Rapport RA-2006-83. Steunpunt Verkeersveiligheid, Diepenbeek.
- Mens en Ruimte (1997). Diepteonderzoek van 120 dodelijke verkeersongevallen in België. Een studie uitgevoerd in opdracht van het BIVV. Eindrapport-Synthese. December 1997.
- Mulder, S., Bloemhoff, A., Harris, S., van Kampen, L.T.B., Schoots, W. (1995). Ongevallen in Nederland. Opnieuw gemeten 1992/1993. Stichting Consument en Veiligheid.
- Scheufele, D. A., & Tewksbury, D. (2007). Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *Journal of Communication*, 57(1), 9-20.
- Shinar, D. (2007). *Traffic Safety and Human Behavior*. Amsterdam: Elsevier.
- Statistics Belgium (2007). Road accident data. Federal Public Service Economy, SME's, Self-employed and Energy.
- Van Hout, K. (2007). Diepteanalyse van verkeersongevallen met motorrijders. Studie aan de hand van verzekeringsdossiers over 145 ongevallen. Eindrapport. Universiteit Hasselt, Instituut voor Mobiliteit, November 2007.
- Verbeek, M. (2004). *A guide to modern econometrics*. John Wiley and Sons Ltd.