Understanding traffic crash under-reporting
linking police and medical records to individual and crash characteristics

Janstrup, Kira Hyldekær; Hels, Tove; Kaplan, Sigal; Sommer, Helle Mølgaard; Lauritsen, Jens

Publication date:
2014

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

● Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
● You may not further distribute the material or use it for any profit-making activity or commercial gain
● You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Understanding traffic crash under-reporting: linking police and medical records to individual and crash characteristics

Kira H. Janstrup\textsuperscript{a}, Tove Hel\textsuperscript{a}, Sigal Kaplan\textsuperscript{a}, Helle M. Sommer\textsuperscript{b}, Jens Lauritsen\textsuperscript{c}

\textsuperscript{a} Department of Transport, Technical University of Denmark, kgs. Lyngby, Denmark
\textsuperscript{b} Department of Applied Mathematics and Computer Science, Technical University of Denmark, kgs. Lyngby, Denmark
\textsuperscript{c} Accident Analysis Group, Orthopaedic Surgery department O, Odense University Hospital, Odense C, Denmark

Abstract

While national crash databases are a valuable resource for the analysis of crash frequency and severity, these databases consist of police reports that are known to be heavily under-reported. This study aligns to the body of research dedicated to estimating the under-reporting rate of crashes by employing the capture-recapture method on available medical and police records. Data consist of records of road users who reported their involvement in a road crash to the police or emergency rooms on the island of Funen in Denmark between 2003 and 2007. Moreover, this study estimates the likelihood for road users reported in police records to appear in hospital records (and vice versa) by estimating joint binary logit models. Results show that the likelihood of appearing in both datasets is positively related to helmet and seat-belt use, number of motor vehicles involved, alcohol involvement, higher speeds and lane number, and females injured. Marital status and education level are not found to be associated with the probability of reporting both to the hospital and the police.

Keywords: Road crash under-reporting; hospital data; police data; capture-recapture; joint binary logit model.

Résumé

Bien que les bases de données nationales de l’accident soient une ressource précieuse pour l’analyse de la fréquence et de la gravité de l’accident, ces bases de données sont constituées de rapports de police qui sont connus pour être fortement sous-déclarée. Cette étude s’aligne sur le corps de la recherche consacrée à l’estimation du taux de sous-déclaration des accidents en employant la méthode de capture-recapture sur les dossiers médicaux et de police disponibles. Les données comprennent des dossiers d’usagers qui ont déclaré leur implication dans un accident de la route à la police ou salles d’urgence sur l’île de Fionie, au Danemark entre 2003 et 2008. En outre, cette étude évalue la probabilité pour les usagers de la route rapportés dans les dossiers de police à apparaître dans les dossiers hospitaliers (et vice versa) par l’estimation de modèles communs logit binaires. Les résultats montrent que la probabilité d’apparaître dans les deux ensembles de données est positivement liée au casque et la ceinture de sécurité, le nombre de véhicules impliqués, la participation de l’alcool, des vitesses plus élevées et le nombre de voie, et les femmes blessés. État matrimonial et le niveau d’éducation ne sont pas trouvés d'être associés à la probabilité de déclarer à la fois à l’hôpital et la police.

Mots-clés: Accident de la route sous-déclaration; données hospitalières, les données de la police, de capture-recapture; modèle joint logit binaire.
1. Introduction

Recent years have witnessed a growing interest in finding methods to estimate the exact number of road crashes. This interest has grown from the fact that police registered crashes are heavily under-reported and, moreover the under-reporting is biased (Derris & Mak, 2007; Elvik & Mysen, 1999). In an earlier meta-analysis of under-reporting studies, Elvik and Mysen (1999) found that the official road crash statistics in Denmark only catch 21% of the hospital injury crashes. In comparison, the under-reporting rates range between 21% and 88% for countries included in the meta-analysis, and between 21% and 57% in European countries. The under-reporting rate varies considerably according to the degree of crash injury severity: Elvik and Mysen (1999) found that the official road crash statistics in Denmark is almost complete when it comes to fatalities with 97% matching the hospital recorded road fatalities. Furthermore, they found that the reporting level for car occupants in Denmark by the police is 48% of that of hospitals, and the same number for cyclists is only 10%. In Europe the reporting rate is generally low for cyclist crashes and ranges between 8% and 66%. Only Great Britain (66%) covers over 30% of the cyclist crashes. In Denmark, Hels and Orozova-Bekkevold (2007) found that in a five year period, only 25% of emergency room cyclist crashes were reported by the police.

In general, two methods have been applied to estimate the total number of road crashes: the capture-recapture method and the method developed by Reurings and Stipdonk (2011). The capture-recapture method is known from ecology, where it is used to estimate the total size of animal populations in the field by repeatedly marking and sampling individuals (Southwood & Henderson, 2000). The capture-recapture method was applied on road crash data from road crashes in the end of the 1990’s by comparing hospital and police crash records and using the crash records that were recorded in both data sources in order to estimate the total number of crashes. The method of Reurings and Stipdonk (2011) is inspired by the capture-recapture method and has been used on data from the Netherlands, although it necessitates access to all emergency room data and not only the ones recorded as a road crash.

The increasing interest in using the capture-recapture method in road safety is reflected in an ample body of literature, accumulated since the 1990’s (e.g., Razzak & Luby, 1998; Morrison & Stone, 2000; Tercero & Andersson, 2004; Meuleners et al., 2006; Amoros et al., 2007; Lateef, 2010; Hassel et al., 2011; Thomas et al., 2012; Tin et al., 2012). Most studies focused on a small subgroup of road users or specific crashes, such as road crashes with work-related vehicles (Thomas et al., 2012), cyclist or pedestrian involved crashes (Tin et al., 2012; Dhillon et al., 2001; Roberts & Scragg, 1994), heavy vehicles involved crashes (Meuleners et al., 2006), road crashes involving children or young people (Roberts & Scragg, 1994; Morrison & Stone, 2000; Dhillon et al., 2001), fatal crashes (Lateef, 2010) or serious injury crashes (Amoros et al., 2007). Only a few studies include all injuries group and all types of road users (Martinez et al., 2012; Tercero & Andersson, 2004; Aptel et al., 1999).

While the capture-recapture method is based on the assumption that some crashes are recorded in both police and hospital data, none of the above mentioned studies attempted to clarify the underlying factors for the crash appearance in the two data sources versus its appearance in a single data source only. In contrast to this, the current study focuses on understanding the under-reporting rate of road crashes in Denmark and revealing the underlying factors for reporting in hospital and police data. The data for the analysis are police and hospital crash records from Funen, the fourth largest island in Denmark, for the period 2003-2007. Almost 10% of the Danish population live on Funen or at one of the nearby small islands, which belong to the emergency rooms on Funen. Therefore, the under-reporting of road crashes in Funen provides a good estimate of road crash under-reporting in Denmark as a whole. The total number of road users involved in a road crash on a part of Denmark, Funen, is estimated for each of the years 2003 to 2007, and the analysed sample comprises road crash data from the Danish Road Directorate and data from all emergency rooms on Funen. A (pseudo) civil registration number for each road user involved in a road crash is listed in the two datasets and the linking procedure is done by those. Besides the traditional variables (e.g. road user type, injury degree, length of hospital stay, time of the day the crash has happened), also a number socio-demographic variables are included in the analysis.

The current study has two main aims. The first aim is to estimate the under-reporting rate of road users involved in road crashes in Funen. The second aim is to estimate the likelihood of reporting in both sources as a function of individual and crash characteristics. Understanding the heterogeneity in the reporting rate of a crash in the two data sources is essential for devising policy measures to increase the reporting rate by targeting specific population/road user groups or specific situational factors. Among the investigated factors are socio-
demographic characteristics, trauma type and severity, and crash characteristics. The applied methodology for estimating the under-reporting rate is the capture-recapture method, while the applied methodology for estimating the likelihood of reporting a crash both to the police and the hospital versus reporting to a single data source is the joint estimation of two binary logit models. The importance of this study lies in the focus on the problem with homogeneity when the capture-recapture method is used. First an estimation of the under-reporting rate is calculated by using the capture-recapture method where the Chapman formulary is used to calculate the total number of road users involved in a road crash at Funen. To understand the likelihood of reporting in the two sources, a function of the individual and crash characteristics is built. The heterogeneity in the reporting rate model is built for data from the emergency room and for data from the police registration.

The remainder of the paper is organized as follows. Section 2 presents the data and describes the variables used in the analysis. Section 3 presents the two methods used for the analysis. Section 4 presents the results, first the under-reporting rate and then the heterogeneity in the reporting rate. Last, section 5 offers a discussion and concluding remarks.

2. Data

There were 27,397 road users who reported an involvement in a road crash to the police or the emergency room at Funen in the years 2003 to 2007. Of these records, 198 were road users who were not registered as Danish citizens and therefore these observations were deleted from the sample which leaves a total of 27,199 road users. 12,637 road users involved in road crashes were registered by the police, and 18,896 road users involved in road crashes were registered by the emergency rooms. Of these crashes, 4,334 were registered in both databases.

The police registered database was obtained from the Road Directorate in Denmark, which collects all information on police registered road crashes in Denmark. The data registered by the police includes crash characteristics, mode types involved, crash location (e.g. intersection, motorway) and collision point(s). Information on the crash circumstances are also listed (e.g. condition of the surface, weather condition, speed limit at the concerned road). Last, some information on the involved parties in the crash is listed (e.g. injury degree of the involved persons, age, gender, civil registration number, municipality). The data registered by the emergency rooms in Funen are collected at three hospitals covering all of Funen (Odense, Svendborg and Middelfart). An AIS (Abbreviated Injury Scale) code is recorded with diagnosis codes related to trauma type, alongside crash characteristics (i.e., number of vehicles involved, the involvement of vulnerable road users and crash location) and personal information of the patients (i.e. age, gender, and civil registration number).

The road user injuries are recorded differently in the two crash registers. The police reports injuries on a four step scale: no injury, slight injury, severe injury, death. The emergency rooms record road user injury on an eight step AIS scale (reference). Thus, to be able to include severity degree in the analysis the end-result of the hospitalization at the emergency room was transformed into an injury scale parallel to that used by the police. If an injured road user’s stay ended with a fatality, the degree of severity was denoted as fatal. If a stay ended with hospitalization, then the degree of severity was denoted as a severe injury, and if a stay ended with the visit at the emergency room or the general practitioner the degree of severity was denoted as a slight injury. Socio-demographic characteristics including education and information on the involved road users’ family were obtained from the database of the Danish Statistical Bureau (Statistics Denmark).

The linking of the three data sets was conducted through the use of the individual civil registration number of the person involved in the road crash. Notably, previous studies (e.g., Meulener et al., 2006; Amoros et al., 2006; Lateef, 2010; Thomas et al., 2012) matched police and hospital records on the basis of matching characteristics (mostly date, gender and age) in the absence of an individual civil registration number. This may lead to false positive identification of matching records when the matched records are highly similar but do not derive from the same crash. The use of individual civil registration number allows to accurately matching of the two data sources without risk of false positive identification of similar crashes as the same crash.

While the whole data set was used to estimate the under-reporting rate, only 26,052 observations of road users involved in crashes were used for understanding the heterogeneity in reporting to the two data sources due to a
high number of records with missing variables. Of these 18,263 records were recorded in the emergency room, 12,062 were police records and 4,273 appear both in the emergency room and the police databases.

3. Methods

3.1. Capture-recapture

A two sampled capture-recapture method is used to estimate the total number of road users involved in a road crash at Funen. The Chapman capture-recapture formulary is defined as follows:

\[ N = \frac{(m+1)(n+1)}{B+1} - 1 \]  

where \( N \) is the total number of road users who report a road crash, \( m \) is the number of road users who have reported exclusively to the police, \( n \) is the number of road users who have reported exclusively to the emergency room and \( B \) is the number of road users who have reported both to the police and the emergency room. The variance and 95% confidence interval (CI) for the estimate of \( N \) is obtained as:

\[ \text{Var}(N) = \frac{(m+1)(n+1)(m-B)(n-B)}{(B+1)^2(B+2)} \]  

\[ 95\% CI = N \pm 1.96 \sqrt{\text{Var}(N)} \]

When using the capture-recapture method, four assumptions have to be met: (i) the population has to be closed, i.e. fixed in number, throughout the estimation period; (ii) there needs to be a perfect and unambiguous identification of subjects common to both registrations; (iii) there has to be independence between the two registrations, meaning that the probability of appearing in one register does not affect the probability of appearing in the other; (iv) there must be homogeneity of capture by a given registration, i.e. all subjects in the set should have the same probability of being registered by one of the two sources. The first assumption means that there should be no entry loss between the two sources, but this is not the case for road crashes, because some road users only get registered by their own doctor, leading to an under-estimation of the total number of injured road users. The second assumption is surely met in this study since the linking is done by an individual pseudo civil registration number. The third assumption is violated when considering a positive relation between appearing in the two sources used in this study, since the police sometimes calls the emergency room and announces them about the road crash. The fourth assumption is difficult to handle and motivates the investigation of the heterogeneity.

3.2. Joint binary model estimation

Two binary logit models are estimated to explore the probability that a crash involved road user is reported in the police register given that the same road user is already reported in the emergency room register and vice versa. The two models are estimated jointly to evaluate whether differences exist when considering the two datasets and to account for differences in the variances of the error terms that would bias a comparison of the estimates. In each binary logit model, the probability of observing a registration match \( i \) for crash \( n \) is expressed by:

\[ P_{ni} = \frac{\exp(V_{ni})}{\exp(V_{ni}) + \exp(V_{nj})} \]

Consider the estimation of the binary logit model \( M \) for the two data sources POL and HOSP: different estimates could result from differences in scale factors, utility parameters, or both. Although parameter estimates \( \beta_{POL} \) and \( \beta_{HOSP} \) are generally considered when presenting model estimation results, the generic estimates correspond actually to the multiplicative terms \( \mu_{POL}\beta_{POL} \) and \( \mu_{HOSP}\beta_{HOSP} \) respectively, since the scale factors \( \mu_{POL} \) and \( \mu_{HOSP} \) are neither identifiable within a particular data source nor separable from the generic utility parameters \( \beta_{POL} \) and \( \beta_{HOSP} \) (Louviere et al., 2000). Accordingly, estimates cannot be directly compared without considering the scale factors, and even data sources generated by the same parameters \( \beta \) with different scale factors would present different estimates \( \mu_{POL}\beta \) and \( \mu_{HOSP}\beta \) (Louviere et al., 2000).
The estimation of the binary logit model from the combination of POL and HOSP provides insight into the differences between the scale factors and enables to perform the likelihood ratio test to evaluate the equality of parameter estimates of the single models (Louviere et al., 2000). The estimation process follows the approach proposed for modelling revealed and stated preference data combination in the context of Logit and Mixed Logit models (e.g., Hensher et al., 1999; Louviere et al., 2000).

The utility function $U_{ni}$ for crash $n$ and registration match $i$ is expressed as follows for data source POL:

$$U_{ni}^{POL} = V_{ni}^{POL} + \varepsilon_{ni}^{POL} = \beta' X_{ni}^{POL} + \alpha' W_{ni}^{POL} + \varepsilon_{ni}^{POL}$$ (5)

where the deterministic part $V_{ni}$ of the utility function contains the vector $X_{ni}^{POL}$ of observable variables common to both data sources, the vector $W_{ni}^{POL}$ of observable variables specific to the POL data, and $\beta$ and $\alpha$ are vectors of parameters to be estimated. The utility function $U_{ni}$ for crash $n$ and registration match $i$ is expressed as follows for data source HOSP:

$$U_{ni}^{HOSP} = V_{ni}^{HOSP} + \varepsilon_{ni}^{HOSP} = \beta' X_{ni}^{HOSP} + \gamma' Z_{ni}^{HOSP} + \varepsilon_{ni}^{HOSP}$$ (6)

where the deterministic part $V_{ni}$ of the utility function contains the vector $X_{ni}^{HOSP}$ of the observable variables common to both data sources, the vector $W_{ni}^{HOSP}$ of observable variables specific to the HOSP data, and $\beta$ and $\gamma$ are vectors of parameters to be estimated. Note that the shared vector $\beta$ implies that trade-offs among attributes included in the vector $X$ are the same for both data sources.

The disturbance terms $\varepsilon_{ni}^{POL}$ and $\varepsilon_{ni}^{HOSP}$ are assumed to be identically and independently Gumbel distributed, their variances representing the different levels of random noise in the data and their ratio $\mu_{POL}/\mu_{HOSP}$ being estimated by normalizing the variance of the data source POL to unity and identifying the relative variance or scale for the remaining data source HOSP. Note that even in this case all the parameters, included the estimated scale parameter, are scaled by the unknown scale parameter $\mu_{HOSP}$.

4. Results

4.1. Results of the capture-recapture method

Table 1 presents the results for the capture-recapture method for the total number of road users involved in a road crash at Funen in each of the years from 2003 through 2007.

Table 1. Number of road users involved in a road crash at Funen by year

<table>
<thead>
<tr>
<th>Year</th>
<th>Matched</th>
<th>Unmatched in police data</th>
<th>Unmatched in emergency room data</th>
<th>Capture (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>927</td>
<td>1,812</td>
<td>2,927</td>
<td>11,381 (10,863-11,900)</td>
</tr>
<tr>
<td>2004</td>
<td>848</td>
<td>1,631</td>
<td>3,033</td>
<td>11,339 (10,729-11,885)</td>
</tr>
<tr>
<td>2005</td>
<td>778</td>
<td>1,499</td>
<td>3,872</td>
<td>10,675 (10,136-11,215)</td>
</tr>
<tr>
<td>2006</td>
<td>876</td>
<td>1,558</td>
<td>2,769</td>
<td>10,122 (9,655-10,589)</td>
</tr>
<tr>
<td>2007</td>
<td>905</td>
<td>1,803</td>
<td>2,961</td>
<td>11,562 (11,024-12,099)</td>
</tr>
</tbody>
</table>

In this number also the material damage only road crashes from the police are included, since some of the road users involved in those actually were registered in the emergency room as well. The number of road users involved in a material damage road crash was in total 7,408. In table 2 the results for the severity degree (fatal, severe and slight) are given and the number of registrations by the police and emergency room are also listed.

Table 2. Number of road users involved in a road crash at Funen, divided by severity degree, transport mode and year

<table>
<thead>
<tr>
<th>Year</th>
<th>Matched</th>
<th>Unmatched in police data</th>
<th>Unmatched in emergency room data</th>
<th>Capture (95% CI)</th>
</tr>
</thead>
</table>

In this number also the material damage only road crashes from the police are included, since some of the road users involved in those actually were registered in the emergency room as well. The number of road users involved in a material damage road crash was in total 7,408. In table 2 the results for the severity degree (fatal, severe and slight) are given and the number of registrations by the police and emergency room are also listed.
The total number of road fatalities at Funen varies from 25 to 39 in the period 2003 to 2007, while the number of severely injured road users varies between 1,190 in 2006 to 1,408 in 2003. The number of slightly injured road users varies from 2,142 in 2006 to 2,317 in 2007. In table 3, the reporting rate for the police in each transport mode can be found.

Table 3. Number of road users involved in a road crash at Funen, divided by severity degree, transport mode, year and the police catch rate

<table>
<thead>
<tr>
<th>Transport mode</th>
<th>Year</th>
<th>Fatal</th>
<th>Severe</th>
<th>Slight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Capture (95% CI)</td>
<td>Police catch rate (%)</td>
<td>Capture (95% CI)</td>
<td>Police catch rate (%)</td>
</tr>
<tr>
<td>Pedestrian</td>
<td>1 (1-1)</td>
<td>100</td>
<td>65 (57-73)</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>2 (2-2)</td>
<td>100</td>
<td>83 (73-93)</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>6 (6-6)</td>
<td>100</td>
<td>50 (46-55)</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>3 (3-3)</td>
<td>100</td>
<td>65 (58-72)</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>7 (5-5)</td>
<td>100</td>
<td>66 (60-73)</td>
<td>62</td>
</tr>
<tr>
<td>Cyclist</td>
<td>6 (6-6)</td>
<td>100</td>
<td>669 (609-739)</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>9 (9-9)</td>
<td>100</td>
<td>637 (688-686)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>4 (4-4)</td>
<td>100</td>
<td>612 (557-668)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>4 (4-4)</td>
<td>100</td>
<td>490 (457-524)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>3 (3-3)</td>
<td>100</td>
<td>582 (522-642)</td>
<td>14</td>
</tr>
<tr>
<td>Moped</td>
<td>4 (4-4)</td>
<td>100</td>
<td>192 (175-209)</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>6 (6-6)</td>
<td>100</td>
<td>212 (190-233)</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>2 (2-2)</td>
<td>100</td>
<td>172 (157-187)</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>6 (6-6)</td>
<td>100</td>
<td>179 (164-195)</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>6 (6-6)</td>
<td>100</td>
<td>201 (186-217)</td>
<td>48</td>
</tr>
<tr>
<td>Motor-Cyclist</td>
<td>2 (2-2)</td>
<td>100</td>
<td>68 (60-76)</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>4 (0-0)</td>
<td>N/A</td>
<td>74 (63-84)</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>0 (0-0)</td>
<td>N/A</td>
<td>70 (58-83)</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>5 (5-5)</td>
<td>100</td>
<td>61 (54-68)</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>3 (3-3)</td>
<td>100</td>
<td>77 (67-87)</td>
<td>44</td>
</tr>
<tr>
<td>Car</td>
<td>15 (15-15)</td>
<td>100</td>
<td>361 (351-371)</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>8 (8-8)</td>
<td>100</td>
<td>291 (285-297)</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>12 (12-12)</td>
<td>100</td>
<td>293 (284-302)</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>10 (9-11)</td>
<td>89</td>
<td>320 (313-326)</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>19 (17-21)</td>
<td>85</td>
<td>349 (340-358)</td>
<td>68</td>
</tr>
<tr>
<td>Bus</td>
<td>0 (0-0)</td>
<td>N/A</td>
<td>14 (6-22)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>0 (0-0)</td>
<td>N/A</td>
<td>10 (10-10)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0 (0-0)</td>
<td>N/A</td>
<td>17 (17-17)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>0 (0-0)</td>
<td>N/A</td>
<td>12 (12-12)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>1 (1-1)</td>
<td>0</td>
<td>11 (11-11)</td>
<td>18</td>
</tr>
<tr>
<td>Other*</td>
<td>1 (1-1)</td>
<td>100</td>
<td>42 (38-45)</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>2 (2-2)</td>
<td>100</td>
<td>38 (36-40)</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>1 (1-1)</td>
<td>100</td>
<td>39 (37-42)</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>4 (4-4)</td>
<td>100</td>
<td>45 (43-47)</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>2 (2-2)</td>
<td>100</td>
<td>31 (29-32)</td>
<td>82</td>
</tr>
</tbody>
</table>

*Include road users of van, tractor and truck.
4.2. Joint model estimation

Table 4 presents the estimation of the joint binary logit models expressing the likelihood that a road user involved in a road crash reported to the police also appears in the hospital record and vice versa.

Similarities exist across the matching likelihood. Gender and age have comparable effects, as male road users have less likelihood to appear in both datasets and increasing age decreases the probability of being recorded in both datasets. Also, higher injury severity makes more likely reporting to both the hospital and the police with respect to minor crashes, and the use of seatbelt and helmets increases the probability of the road users appearing in both datasets. Last, morning peak crashes are more likely to be matched across datasets and comparable seasonal effects can be observed, as the reporting from police to hospital and from hospital to police seems more likely in the summer than in the colder seasons.

Interestingly, differences exist across the matching likelihood, not only because of the presence of variables specific to the HOSP and POL datasets. When considering being in a car as reference category, pedestrians are more likely to appear in both datasets regardless of analysing the reporting in the hospital or the police, and cyclists are more probable not to appear in the other dataset. However, moped and motorcyclists are underreported with respect to car occupants when checking whether hospital records are present in the police ones, and overreported with respect to car occupants when controlling whether police records are present in the hospital ones. When considering the number of parties involved in the crash, both models express the same tendency, but to a different extent as the likelihood of police records appearing in the hospital ones appears higher. In the hospital records, with respect to the reference category of injuries to the upper extremities, injuries to the head, the thorax and the spine are more likely to be related to an increase in the probability of the crashes being also in the police records. The same correlation is observed with combinations of injuries, in particular when spinal injuries are recorded. In the police records, there is not a significant effect of the level of education of the road user in reporting the crash also to the hospital. A higher probability of reporting the police record also to the hospital is correlated with the road where the crash occurred being larger and having a higher speed limit.

The estimation of the joint models allowed estimating the scale parameter $\mu_{\textit{POL}}$ with respect to the normalized scale parameter $\mu_{\textit{HOSP}}$. The estimate significantly lower than 1 indicates that the variance of the error term $\varepsilon_{\textit{POL}}$ is larger than the variance of the error term $\varepsilon_{\textit{HOSP}}$, which in turns indicates that the police dataset contains more noise than the hospital dataset.

Table 4. Estimates of the joint model of the likelihood that a road user involved in a road crash reported to the hospital appears in the police records and vice versa

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>HOSP Estimate</th>
<th>t-stat</th>
<th>POL Estimate</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>-0.603</td>
<td>-14.23</td>
<td>-0.895</td>
<td>-16.88</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>-</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Less than 18 years old</td>
<td>-0.872</td>
<td>-13.12</td>
<td>-0.755</td>
<td>-9.10</td>
</tr>
<tr>
<td></td>
<td>18-24 years old</td>
<td>-0.916</td>
<td>-13.18</td>
<td>-0.941</td>
<td>-10.72</td>
</tr>
<tr>
<td></td>
<td>25-34 years old</td>
<td>-0.910</td>
<td>-13.07</td>
<td>-0.979</td>
<td>-11.03</td>
</tr>
<tr>
<td></td>
<td>35-44 years old</td>
<td>-0.855</td>
<td>-11.35</td>
<td>-1.010</td>
<td>-10.59</td>
</tr>
<tr>
<td></td>
<td>45-54 years old</td>
<td>-0.854</td>
<td>-10.32</td>
<td>-0.979</td>
<td>-9.53</td>
</tr>
<tr>
<td></td>
<td>55-64 years old</td>
<td>-0.723</td>
<td>-7.05</td>
<td>-0.740</td>
<td>-5.85</td>
</tr>
<tr>
<td></td>
<td>65-74 years old</td>
<td>-0.831</td>
<td>-7.17</td>
<td>-0.623</td>
<td>-4.44</td>
</tr>
<tr>
<td></td>
<td>Over 75 years old</td>
<td>-</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Injury severity</td>
<td>Minor</td>
<td>-</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serious</td>
<td>1.600</td>
<td>30.64</td>
<td>2.510</td>
<td>25.20</td>
</tr>
<tr>
<td></td>
<td>Fatal</td>
<td>5.210</td>
<td>9.72</td>
<td>2.740</td>
<td>10.28</td>
</tr>
<tr>
<td>Road user type</td>
<td>Pedestrian</td>
<td>1.270</td>
<td>11.53</td>
<td>0.556</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>Cyclist</td>
<td>-0.345</td>
<td>-5.00</td>
<td>-0.546</td>
<td>-5.66</td>
</tr>
</tbody>
</table>
Moped 0.870  8.74  -0.883 -6.64
Motorcyclists 0.263  1.84  -1.280 -6.63
Car - - - -
Van 0.889  4.59  0.819  3.91
Heavy vehicle -0.999  -6.02  -0.987  -5.38

Seatbelt
Yes 2.170  34.60  2.090  26.75
No - - - -

Helmet
Yes 0.515  6.23  0.315  3.03
No - - - -

Family status
Single -2.650  -35.95  -2.430  -26.91
Partner -2.930  -42.02  -2.730  -31.42
Other status - - - -

Other parties involved
Zero - - - -
One 1.070  18.01  0.253  3.29
Two 0.863  11.43  0.315  3.03
Three or more 1.010  12.40  0.572  5.30

Type of injury
Head 1.110  19.08  - -
Head and thorax 1.600  9.20  - -
Head and upper extremities 1.710  16.92  - -
Head and lower extremities 1.980  18.09  - -
Head and spine 2.710  16.79  - -
Thorax 1.630  13.56  - -
Thorax and upper extremities 1.480  7.39  - -
Thorax and lower extremities 2.110  11.80  - -
Thorax and spine 2.220  9.36  - -
Upper extremities - - - -
Upper extremities and spine 2.690  11.39  - -
Lower extremities 0.880  13.12  - -
Lower extremities and spine 3.120  12.86  - -
Spine 2.240  19.35  - -

Education
Low education - - - -
Medium education - - 0.050  0.34
High education - - -0.008  -0.13

Speed limit
Less than 70 km/h - - - -
70-90 km/h - - 1.160  17.11
100-130 km/h - - 1.370  9.84

Number of lanes
One - - - -
Two - - 3.510  36.96
Three or more - - 3.510  29.57

Type of day
Weekend - - -0.085  -1.62
Weekday - - - -

Time of day
Morning peak 0.107  1.78  0.211  3.02
Other periods - - - -

Season
Spring -0.744  -13.10  -0.614  -9.22
Summer - - - -
Autumn -0.703  -13.00  -0.606  -9.45
Winter -0.739  -12.77  -0.539  -8.01

Scale parameter 1.000  -  0.886  -7.73*
Log-likelihood at zero -33315.148
Log-likelihood at convergence -15639.149
Adjusted Rho-bar squared 0.528

Note: * t-test with respect to 1 (tests the equality of the scale parameters)

5. Conclusion & Discussion

In this paper, we estimated the total number of road users involved in road crashes on the basis of the capture-recapture method. We also estimated joint binary logit models to identify the likelihood that a road user involved in a road crash and reported to the hospital will also be in the police records, and vice versa. The total number of road users involved in a road crash at Funen in the study period turned out to be much higher than the police recorded number. The number of fatalities was different in the two registration sources; this could be due to the fact that presumed suicides and sudden diseases before the crash (e.g. stroke) are excluded from the police records (see Pompili et al., 2012). The missing registration of fatalities at the emergency room could be because road users who have died in a road crash are taken directly to the mortuary and therefore do not get registered in the emergency room. Also, the different numbers of recorded fatalities could be due to the fact that, in some
cases, the police simply does not know about the crash. As expected, many vulnerable road users with slight injuries or even severe injuries from road crashes only report to the emergency room. In the cases that slightly injured road users only got registered to the emergency room, it could be due to the fact that if a road user has injured the arm or hand in a road crash and nothing else has happened, they turn to the emergency room only and do not find it necessary to involve the police. The number of injured bus road users are also highly under-reported, possibly because there are many road users injured in these types of road crashes and the police does not have time to collect info on all road users. The reporting rate for injured road user registered by the police is a bit lower than the numbers found in other studies (Amoros et al., 2007; Aptel et al., 1999), possibly because the police in Denmark only has to report the crash if the police officer at the crash scene think that this road crash is serious enough or if the road user wants to use the report for an insurance case.

Model estimates showed significant correlation of gender and age, with lower reporting among males and among older road users. This finding could relate to the fact that children are always passengers that need help to get out of the car and are always taken to a health check at the emergency room after involvement in a road crash. The gender difference could be due to the fact that females in general are more aware of their own health. Crashes involving vulnerable road users are often more severely hurt, but their reporting is different across datasets: with respect to car drivers, pedestrian are more likely and cyclists are less likely to appear in both sources, while moped and motorcyclists in police records are less likely to appear in hospital ones but the opposite is recorded when the comparison is reversed. Crashes involving seat belt and helmet use were found for both models to have a higher probability of getting reported in both sources, possibly because seatbelt and helmet use indicates awareness. The number of motor vehicles involved in a road crash turned out to be positively correlated with the probability of getting reported in both sources, likely because of the number of people who are involved in the crash. Road users involved in crashes on roads with higher speed limits and higher number of lanes had a higher probability of getting reported in both sources, possibly in relation to the “visibility” of the crash. As expected, the severity degree of road users involved in a road crash was found highly significant in both analysis and the probability of reporting to the police and the emergency room increased with the severity degree. Often the police arrives first to the crash site, and in severe cases calls an ambulance and decides which road user should be sent to the emergency room. That explains for example the fact that road users who have head, thorax, and spine injuries are more likely to have been reported in both databases, since the police may have sent road users with these injuries to the hospital while arm injuries appear less severe and less urgent to treat. Last, the estimate significantly lower than 1 indicates that the variance of the error term $\varepsilon_{HOSP}$ is larger than the variance of the error term $\varepsilon_{POL}$, which in turns indicates that the police dataset contains more noise than the hospital dataset.

Overall it appears that a loss of information about the road users involved in road crashes is lost when only police recorded road crashes are included in crash modelling and that the number of severe and slight injuries are highly under-reported by the police. To get a more correct picture of the amount of road users involved in a road crash in Denmark it is necessary to include other registration sources as well, as in this study considering emergency room data.

Acknowledgements
The financial support of the Danish Council for Strategic Research to the project “Improving Road Safety” is gratefully acknowledged.

References


