Denne artikel er publiceret i det elektroniske tidsskrift
Artikler fra Trafikdage på Aalborg Universitet
(Proceedings from the Annual Transport Conference at Aalborg University)
ISSN 1603-9696
www.trafikdage.dk/artikelarkiv

A case study on concordance between self-reported accidents and records by hospital and police

Ph.d.-student Katrine Meltofte (krm@civil.aau.dk), Aalborg University
Ph.d.-student Tanja K. O. Madsen (tkom@civil.aau.dk), Aalborg University
Assistant Professor Anne Vingaard Olesen (avo@civil.aau.dk), Aalborg University
Associate Professor Harry Lahrmann (hsl@civil.aau.dk), Aalborg University

Abstract

A self-report questionnaire on bicycle accidents was distributed to 6.793 respondents every month during one year (2012/2013). This paper evaluates respondents' understanding of the questionnaire's phrasings and content, and provides insight in the concordance between hospital data, police data and self-reports concerning the number of accidents as well as basic accident information.

694 accidents were reported by respondents. 99 of these were self-reported to have had medical care at the hospital; 74 of these accidents could be matched to a hospital record. Only 53 were actually registered in the hospital records as having had a traffic accident. Information on the primary counterpart is compared, and good agreement is found between self-reports and hospital records (Kappa =0.63).

23 respondents self-reported that a police report had been filed, yet only 6 accidents could be located in the police files.

The findings are relevant to the level of underreporting of bicycle accidents in Denmark, as it indicates that there could be substantially more accidents that noted in official records.

1. Introduction

Many studies have been made in the field of underreporting of traffic accidents. Most often the level of underreporting is researched by comparing data from hospital records and police records in order to estimate the number of accidents that are missing in police records as seen in for instance in meta-analysis (Elvik, Mysen 1999) or other studies (Bull, Roberts 1973), (Broughton, Keigan et al. 2010) or (Boufous, Ivers et al. 2010). A small number of studies debates the police underreporting by comparing with other data sources, such as insurance data or self-reported accidents, for instance (Isaksson-Hellman 2012) and (Finestone, Guo

et al. 2011). But no matter which data sources one compare to estimate validity of accident numbers or information the same challenges arise; we do not know the true number of accidents and we do not know all about the flaws of the collected accident data. As police reported accidents have been the primary source of accident information in many countries, the literature debating validity of police reports is quite extensive, but hospital data and self-reports of accidents are two data sources whose validity are seldom researched. But the fact is that all data sources are flawed; self-reports, police records (Austin 1994, Austin 1993) and hospital records as well.

Self-reports are sometimes distrusted in traffic research due to several reasons. Forgetting and/or memory loss may influence the correctness of recall (Lajunen, Özkan 2011) as well as social desirability effects (Wåhlberg, Dorn et al. 2010, Wåhlberg 2010) . Self-reports can also be suspected to suffer from statistical bias due to under-reporting by those with many crashes and possible over-reporting by some subgroups (Tivesten, Wiberg 2013, Wåhlberg 2009). The agreement between self-reported accident data and other data sources are sometimes low, which can be mentioned as a problem with the self-reported data (Wåhlberg 2009). But low agreement with other data sources does not in itself diminish the validity of self-reported data; it depends on the validity of the data source to which the self-reported data are compared.

Police records are also flawed. A study by Danish researchers estimates that only approximately 14 % of severe accidents on bike registered at hospitals in Denmark are also registered by the police (Janstrup, Hels et al. 2014). The large discrepancy between numbers of traffic accidents reported by the police and the number recorded by the hospital is a fine indication that the police records lack information on many accidents. As the discrepancy between hospital records and police records among other things varies with injury severity and mode of transport (Janstrup, Hels et al. 2014) the police records are biased.

Very little is known about the flaws and bias in hospital records. No studies on validity and reliability of hospital-recorded accidents have been found, however, as with all records, reporting errors could theoretically be present. Thus, errors could arise from e.g. personnel misfiling records or patients not telling medical personnel that their injuries were sustained in traffic. If subgroups choose not to seek medical care, this would also lead to bias in medical records.

It is obvious that none of the above mentioned data sources are complete or flawless, but must be used with the possible errors in mind. The current study contributes to this field of study by examining aspects of the validity of the self-reported accident data and hospital records, while also commenting on some methodological issues regarding self-report questionnaires.

The following hypothesizes are stated:

A) Self-reporters understand the phrasing of the questionnaire and can locate their accident on Google Maps.
B) Self-reporters who claims to have been in contact with either police or hospital personnel in connection with their accident can be located in police or hospital records
C) The self-reported accident information is congruent with information in the records written by hospital and or police personnel.

2. Method

2.1. Data sources and definitions

Each month the participants were contacted via email in which there was a link to an online questionnaire. In the monthly questionnaires, the respondents were asked whether or not they had had an accident in the previous month. If they answered that an accident had taken place, they were then given more questions on the nature of their accident. A simplified form of the information asked in the questionnaire can be seen in Table 1.

Table 1: The design of the monthly questionnaire in comparison with police and hospital records. Only data from police and hospital records that are comparable with the self-reports are shown in the table; police and hospital records contains more information than shown below, for instance alcohol consumption. [M] = Multiple choice, [D]= Descriptive, [Opt]=Optional, not necessary recorded, [A]= Only recorded if Activity is filed as a traffic accident

Topic	Self-reported information	Police-records	Hospital-records
Time and date	Date [M] Month[M] Year[M] Time (hourly interval) [M]	Date[M] Month[M] Year[M] Time (exact) [M]	
Place	Location - coordinates from maps.google.dk Place[D]	Location – distance from a selected point	Place [M]
Circumstances	Road surface[M] Visibility[M] Day/Night[M] Weather[M] Street light [M]	Road surface [M] Visibility[M] Day/Night[M] Weather[M] Street light [M]	Road surface [Opt] Activity [M]
Safety measures	Use of lights on bike [M] Use of fluorescent bike jacket [M] Use of helmet[M]	Use of lights on bike [M] Use of helmet[M]	Use of helmet [Opt]
Accident	Other parties or solo accident[M] How many parties[M] Primary counterpart[M] Accident description [D]	Other parties or solo accident[M] How many parties[M] Primary counterpart[M] Accident description [D]	Other parties or solo accident[M, A] Primary counterpart[M, A]
Severity	Most severe injury[M] Injuries[D] Medical treatment[M] Absenteeism[M] Police report[M] Insurance report[M]	Most severe injury[M] Medical treatment[M]	Injuries[M] Medical treatment[M]

The questionnaire is designed so that many of the questions are comparable with those found in police reports; this makes it possible to easily compare information from the two data sources with regards to accordance of information.

As can be seen in the table, police records contain varies types of accident-information, both as multiple choice data and descriptive fields, and all fields must be filled out in the police report before it is filed. Not all accidents that come to the knowledge of the police have to be registered in the police reports. Only accidents where at least one of the following circumstances are present (Hemdorff, Lund et al. 2003):

- There is a person injured (bruises and abrasions does not constitute as injuries)
- Material damage of more than 50.000 DKK per vehicle or 5.000 DKK per other equipment
- Foreigners are involved and there has been made a compensation claim against them
- A person employed at the police is involved in the accident
- There has been a violation of the Road Traffic Act that should lead to preliminary charges

In contrast to this, the hospital records all people, who seeks medical care – no matter the severity of their injuries or the extent of material damage. In the self-report questionnaires respondents could state all accidents and near-misses that they had encountered no matter their physical consequences. This meant that people also answered the questionnaire telling about minor accidents with little or no personal injury or material damage. Thus there are differences in the three types of data sources with regards to what accidents one could expect to find recorded. Hospital records would expectedly contain mostly accidents where people was injured (hence their need to seek medical care), police records would expectedly contain accidents with more serious injuries and self-reports would be expected to contain severe accidents as well as accidents with no injuries or even near-misses.

As well as differences in severity of the expected records in the three data sources, they also use different definitions of what constitutes as an accident. The police use the following definition of a traffic accident (Hemdorff, Lund et al. 2003):

- The accident must have happened on public road/square/area and at least one of the involved parties was driving.

The self-report survey defines a traffic accident as an event that has happened in Denmark and on public road, where the self-reporter was travelling by bike and at least one of the following happened (Lahrmann, Madsen et al. 2014):

- The bicyclist had physical contact with a counterpart
- The bicyclist fell over, was hurt or experienced material damage because of a counterpart's behavior. This also includes events where no physical contact was present.
- The bicyclist fell over or was hurt while biking, even though there were no other parties involved.

The hospital records define a traffic accident as an accident that involves means of transport or an animal, that are used to transport persons or goods from one place to another (Sundhedsstyrelsen 2009). This means that sporting activities will not be registered as traffic accidents. But in contrast to police records all modes of transport, including pedestrians, will be registered regardless of the accident was a solo or multiple party accident.

The main differences in the three sets of data regarding their reporting constraints thus seems to be that police records does not have information of accidents that have happened outside public road and only

records of people with injuries or material damage exceeding 5.000 DKK. The self-reports contains all events that have happened while biking; also where no injuries was sustained or no material damage done. The hospital records should contain records of the persons who, when asked at the emergency room, tell that they have been injured while using their bike as a mode of transport.

As seen in Table 1, the difference between hospital records and police records or self-reports is obvious; only very little accident-related information is present in hospital records as a standard. To understand the difficulties of comparing information from hospital records with self-reported data and police data, it is necessary to understand the differences in how information is presented and recorded in LPR. As an exception are the data in LPR that are recorded at the emergency room in Odense on the Danish island of Funen, where an extended recording is carried out. But in the rest of Denmark, only very basic information on accidents are required to be inputted in the hospital records: The action (for instance falling), the injury mechanism (for instance hitting one's head), the place of the accident (for instance public road), activity (for instance transport to/from work). If the activity is chosen to be one that constitutes as a traffic accident, the hospital personnel must also register the mode of transport for the injured (for instance bike) and the mode of transport for the counterpart (for instance pedestrian). Yet, when scrutinizing the data, one finds that the entries in hospital records vary in detail; it seems the some medical personnel register more information in the hospital records than the minimal requirements.

2.2 Self-reported accidents – participants and questionnaire design

The self-reported accident questionnaire has been developed and distributed as part of a project on the traffic safety effects of wearing a brightly colored jacket while bicycling (Lahrmann, Madsen et al. 2014 - submitted to journal), (Lahrmann, Madsen et al. 2014). 6.793 participants was part of the experiment for the duration of one year, beginning in November 2012 and ending in October 2013. A group of test subjects (N = 3.402) were given a bright yellow bike jacket (Lahrmann, Madsen et al. 2014); for a duration of 12 months the persons were asked to use the jacket when biking and to answer a questionnaire each month. The questionnaire consisted of questions regarding their use of the jacket and the participants' potential accident involvement. Besides the test subjects, there was also a control group (N= 3.391), where the participants was asked to fill out a questionnaire on any occurring accidents each month. The control group was given the yellow bike jacket after the experiment ended. The project's original purpose was to estimate the effects on bicyclist safety when wearing brightly colored clothes, but in this paper the data from the monthly questionnaires will be used for research on self-report questionnaires and the accordance of answers with other data sources such as hospital records and police records.

2.2.1 Localizing on map

When the respondents reported an accident in the monthly questionnaires, one of the questionnaires was to localize the accident. To do so they were asked to open a new tab in their browser, localize the place of their accident on Google Maps and copy-paste the coordinates into the questionnaire. The participants were also given the opportunity of explaining the location of the accident in a text field instead of copy-pasting coordinates – they would then be contacted via telephone by a member of the project crew to help them find the correct coordinates together.

2.2.2 Phrasing questions on the number of involved "parties" [1]

One of the difficulties when it comes to designing a self-report questionnaire is how to gain detailed information on exactly how the accident happened. Due to technical reasons it was not possible for respondents to draw or make sketches of their accident, and accident information thus relied only on multiple choice answers or information given in descriptive fields. The multiple choice question was phrased: *"Besides yourself, where there any other parties involved in the accident?"*[2] If the respondent answered "*Yes*", they then got the question: "*You have informed that there were other parties involved in the accident. How many parties (not persons) was involved in the accident besides yourself (a part could be a car, a truck, a bicyclist, a pedestrian, an animal and alike)*?"[3] The possible answers were: "*1 party in addition to myself* "[4],"*2 parties in addition to myself*"[5] " and "*More than 2 parties in addition to myself*"[6]". On a dropdown menu, the respondent was then asked to choose the answer to the question: *"Who was your primary counterpart in the accident*"[7]. These multiple choice answers were then compared with the information the respondent had given in the descriptive text field, as we suspected that there could be inconsistency between the descriptive text and the answers of the multiple choice question because the *"number of parties"* could be considered a technical term. For a non-professional this could be interpreted as "*How many persons were involved in the accidents*", which could lead to a very different number of parties than if one used the definition most often used in accident analysis, where "party" refers to the number of traffic elements involved in the accident (eg. number of vehicles and/or pedestrians).

To test if the respondents understood the phrasing, we manually compared all answers of the multiple choice question with the description of the accident and the circumstances leading to the accident that the respondent had written.

Matching records

To test the hypothesizes of accordance of self-reported information with official accident data, three different data sets are used: The self-reported accident information, on which more information is given in the following sections, and data from the police accident database and from the National Patient Register[8] (hereafter denoted as LPR) that contains the hospital records.

All Danish citizens and people with a residence permit are given a unique social security number (herafter denoted as a CPR-number) that are used as an identifier in LPR and the police accident database. Hence, it is possible to match recordings from the two databases and the self-reported accident information as long as the respondents provide their CPR-number in the self-report questionnaire.

Thus, a match is based on agreement of CPR-number. Time constraints must also be taken into account as a respondent could have an accident after having self-reported another. To minimize the risk of matching a

[1] In Denmark, it could be considered a technical term to ask how many parties were involved in an accident. The original phrase in Danish is: "part", and is not necessarily considered a laymans term.

[2] Authors translation – original question in Danish: "*Var der andre parter, ud over dig selv, impliceret i uheldet?"*

[3] Authors translation – original question in Danish: *"Du har oplyst, at der var andre parter involveret i uheldet. Hvor mange parter (ikke personer) var udover dig selv impliceret i uheldet (en part kan være en bil, en lastbil, en cyklist, en fodgænger, et dyr eller lignende)?"*

[4] Authors translation – original answer in Danish: *"1 part udover mig selv".*

[5] Authors translation – original answer in Danish: *"2 parter udover mig selv"*

[6] Authors translation – original answer in Danish: *"Flere end 2 parter udover mig selv"*

[7] Authors translation – original question in Danish: *"Hvem var din primære modpart i uheldet?"*

[8] *Landspatientregisteret* in Danish

self-reported accident with an accident occurring later, we only consider the records to match if the date of the hospital admittance is between on month prior to the self-reported accident date and two months after. As respondents are asked to self-report every month it would be highly unlikely that they forget their accident date and state the accident date as more than one month wrong. To ensure that respondents, who did not seek medical attention immediately after their accident, is not excluded from the match, we consider accidents that have been recorded by the hospital up to two months after the self-reported accident date a match. There have been found no studies on how much time people wait to seek medical care due to an accident that could be used to verify the choice of time constraints in matching.

If it is obviously stated in LPR that the record is not correlated to a traffic accident, the record will not be matched – for instance if the records show the injuries are burn marks and the accident happened indoors at home. When we cannot exclude the possibility that the injuries could be sustained from a traffic accident, the records are considered a match.

3. Results

The results on the respondents' abilities to fill out the questionnaire will be addressed first, as the first hypothesis was:

A) Self-reporters understand the phrasing of the questionnaire and can locate their accident on Google Maps.

Secondly the result on matching self-reports and hospital records and police records will be presented to shed light on the last two hypothesizes:

B) Self-reporters who claims to have been in contact with either police or hospital personnel in connection with their accident can be located in police or hospital records
C) The self-reported accident information is congruent with information in the records written by hospital and or police personnel.

3.1 Respondents' ability to fill out the questionnaire correctly

The participants (both test and control group) reported 833 bicycle accidents during the test period of one year. Of these there were 590 accident reports where the location was copy-pasted correctly from Google Maps (71%). "Correctly" meaning, that the respondent managed to input coordinates in the format xx.xxxxxx, xx.xxxxxx. Yet, it must be emphasized that it is unknown whether or not the copy-pasted coordinates are correct.

19 respondents copy-pasted the location from Google Maps, but in a format that did not correspond with what they were asked (eg +xx° xx' x.xx", +xx° xx' xx.xx" or the correct format of coordinates followed by an address).

224 respondents did not input a correct value; they either left the field blank, entered an address or copy-pasted the URL to google.maps.dk.

Out of the 833 accident reports, only 47 (6%) had to be corrected manually because the respondents chose a wrong party. This was due to the respondent either writing a number of parties or choosing a primary counterpart that did not coincide with his/hers accident description.

3.2 Self-reported contact to hospital, police and insurance

Of the 833 filed self-reports from the participants only 694 reports can be used in a study of the participant's accident involvement. The remaining 139 reports are either near-misses (58 reports) or accidents that do not meet the study criteria (81 reports in total). Not meeting the criteria could be e.g. accidents that took take place foreign countries, accidents when using another mode of transportation than bicycling or accidents that did not happen on public road (often related to mountain biking).

As can be seen from Table 2, the hospital is the most frequently contacted closely followed by the insurance company. Police contact and a visit to the General Practitioner (hereafter denoted GP) are much more uncommon.

Table 2: Where self-reporters claim to have made contact due to their accident. N = 694
(*)The self-reporters did not specifically answer no to this, as the question of seeking medical attention had another phrasing.

	Yes	No	Do not know
Police	23	667	4
Insurance	96	585	13
Hospital / Emergency room	99	595*	0
Only GP and not hospital	21	673*	0

The number who claims to have made contact with their insurance company is quite high which indicates that a way of gaining knowledge on traffic accidents could be to examine insurance company records. But as can be deducted from Figure 1, 57% of all who made an insurance claim also stated that they needed medical attention at either the hospital or at the emergency room. Basing accident research on the hospital records in this case seems to yield data from more than half of the accidents which were also reported to the insurance company.

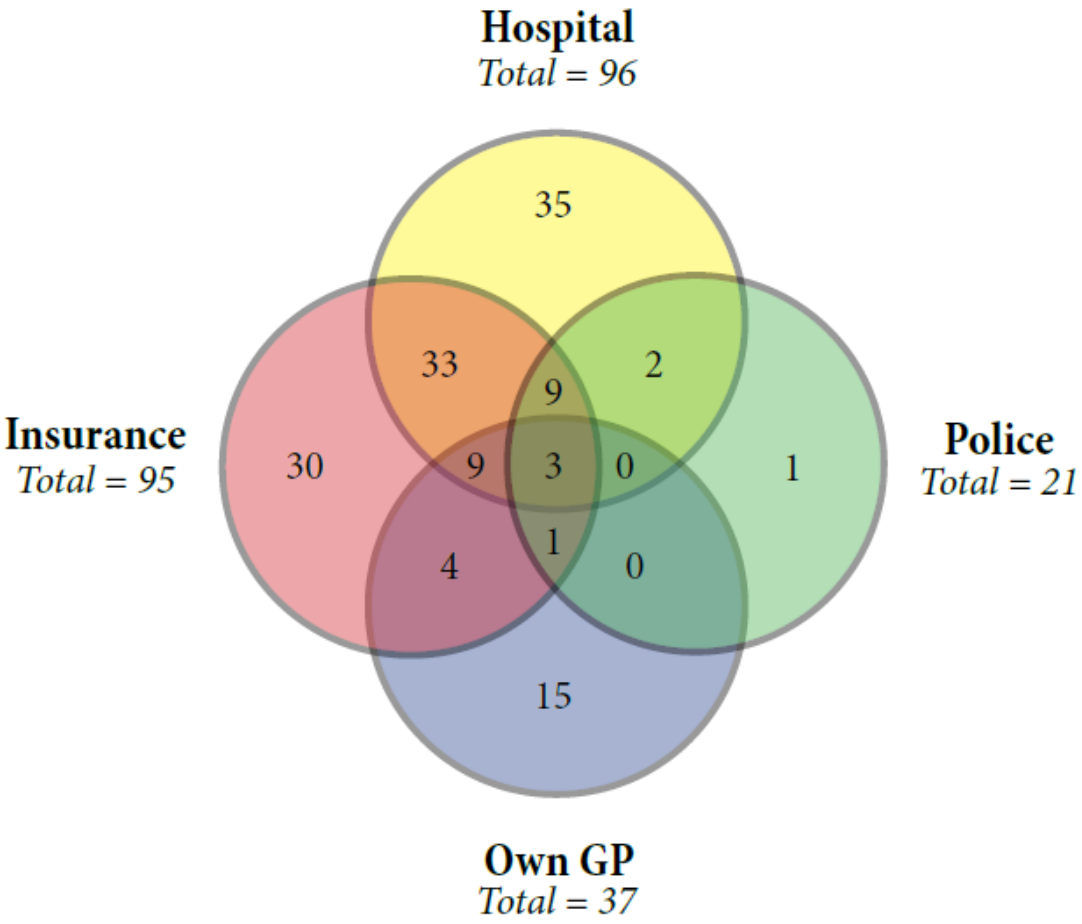

Figure 1: An Euler diagram of the self-reported data. Overlap in police and insurance only = 6, overlap in own GP and hospital only = 5. Answers of "do not know" are excluded from the diagram; this is the reason for the disagreement between total numbers and summation of the numbers shown in the Euler-diagram.

The CPR-number of respondents, with the 694 self-reported accidents that meet the study criteria and are not considered near-misses, are compared with records in LPR and the police database. The rules for matching as described in the *"Methods"*-section yields 72 good matches between respondents' self-reported accidents and accidents registered in LPR. 2 records match within the time constraints, but have such a vague description in LPR that it is not possible to exclude the possibility that the injuries was sustained by an accident and not by something else. We include them in the category of matched records as it is unlikely that persons claiming in their self-report questionnaire that they had been to the hospital due to a traffic accident was in fact not, but instead within days from the accident was in the emergency room with injuries sustained by for instance "other activities" in "a residential area".

The distribution of recorded activities leading to the accident can be seen in

Table 3.

Table 3: The recorded activities leading to injuries registered in LPR for all respondents that are considered to have a match between hospital records and self-reported accidents.

Activity	Code	Number of accidents	Considered a traffic accident in LPR?
Paid transportation work	EUA0	1	Yes
Transport between workplaces	EUA02	2	
Paid work	EUA1	1	No
Paid work, unspecified	EUA19	1	
Transportation in spare time	EUA2	39	Yes
Transport to/from work	EUA21	1	
Other unspecified transportation	EUA28	1	
Transportation, unspecified	EUA29	9	
House activities and unpaid work	EUA3	1	No
Sports during school/education	EUA42	2	No
Sport and exercise	EUA5	8	No
Vital activity, unspecified	EUA79	1	No
Other activity	EUA8	5	No
Activity not specified	EUA9	4	No
Total		74	*Yes= 53* *No= 21*

The same time constraints are applied to the matching of self-reports and police records. For a match to be accepted, the respondents in the police accident database must also be registered as injured in traffic while riding a bike, as the self-reports concern bike accidents. The rules for matching yields 6 recorded accidents in police files.

To see if the percentage of accidents reported in hospital and police databases deviates from the general norm in Denmark, a comparison with official statistics has been made. The current study finds that when 53 respondents are found to be registered in LPR with a traffic accident, 6 are to be found in the police database.

A comparison between the reporting level from this study and the reporting level of all bicycle accidents in Denmark in 2012 and 2013 can be seen in Table 4. A two-tailed Z-test is applied and significant ($p<0.05$) difference is found between the reporting levels ($p=0.034$).

Table 4: The number of accidents with bicyclists (solo accidents and multiple parties) in police records and hospital records. Data from 2012 and 2013 (Danmarks Statistik 2014). The level of reporting by police is calculated as the number of police records divided by number of hospital records multiplied by 100.

	2012 in total	2013 in total	Mean 2012-2013	Current study
Police records	826	796	811	6
Hospital records	16670	15991	16331	53
Level of reporting by police	4,96%	4,98%	4,97 %	11,32%

3.4 Accordance between LPR data and self-reports

The validity of self-reports have often been discussed with the use of another data sample to compare with; the level of agreement between the two datasets are then considered to indicate the validity of the self-reports as the official data from police or hospital records most often are considered as more reliable, (Arthur, Tubre et al. 2001). It is difficult to compare self-reported data with hospital records to discuss validity as the hospital does not record that many data on accidents. For instance, it is not necessarily stated in the hospital report whether or not the bicyclist was using helmet or if he/she was riding on a bicycle lane. But if the injury is registered with an activity of either "paid transportation work" or "transportation", the hospital also have to register the counterpart involved in the accident (Sundhedsstyrelsen 2009). Thus it is possible to compare the counterpart that self-reporters gave in the questionnaire and the counterpart that is stated in the hospital records. Table 5 shows the counterparts stated in hospital records and the counterparts from the self-reported questionnaires.

*Table 5: The accordance between self-reported counterpart and the counterpart entered in hospital database. N= 53 , *)The counterpart cannot be registered by EUM-codes*

Counterparts	LPR [number of cases]	Self-registered [number of cases]
No counterpart (EUM0 or EUU0)	32	24
Pedestrian (EUM1)	3	2
Bicyclist (EUM2)	4	10
Moped (EUM3)	1	1
Motorcycle (EUM4)	2	1

Car (EUM5)	10	13
Not specified (EUM9), or other (Truck/bus, animal) *	1	2

It can be seen in Table 5, that there are some differences between official hospital records and self-reports with regards to the counterpart. It seems that more solo accidents are registered in the hospital files and fewer bicyclists as counterparts. Cohen's Kappa is calculated for the concordance between counterpart registered in LPR and registered in self-reports. A Kappa value of 0.63 is found, which is interpreted as good agreement between the datasets.

Table 6: The self-reported information of having medical care, filed police report or made insurance claim

		All matched [N= 74]	**Matched and registered in LPR with a traffic accident** [N=53]
Medical treatment	Yes	63	49
	No	11	4
Police report	Yes	11	9
	No	61	42
	Does not know	2	2
Insurance claim	Yes	40	31
	No	32	20
	Does not know	2	2

In Table 6 the self-reported information of contact with hospital, police and insurance company can be seen. Interestingly, 11 of the 74 respondents, who were matched to a record in the hospital database, have answered no to the question of having medical care. This could be either due to poor matching, respondents forgetting the medical examination, respondents not believing that the examination constitutes as "medical care" or simply that the respondents sought medical care days after they have finished the questionnaire.

3.5 Accordance between police records and self-reported information

A comparison of information present in self-reports of accident and police records are given in Table 7; it can be seen that almost all information is congruent.

Table 7: A comparison of information given in self-report questionnaires and registered in police records of the same accident. The number indicates how many of the 6 records contain the given information.

	Self-reports	Police records
Weather	6: No rain	6: No rain
Road surface	6: Dry	5: Dry 1: Wet
Daylight	4: Daylight 1: Night 1: Does not remember	5: Daylight 1: Night
Lighting	1: No street lights present 1: Street lights on 3: Street lights not on 1: Does not remember	1: No street lights present 1: Street lights on 4: Street lights not on
Visibility	5: Good visibility 1: Does not remember	6: Good visibility
Helmet	5: Using helmet 1: Not using helmet	5: Using helmet 1: Not using helmet
Injuries	1: Femoral fracture 1: Concussion 1: Bruises and similar 3: Other	2: Fracture or ligament injury in hip/leg/foot 2: Slight injuries or admitted for observation (concussion) 1: Lesion on spine/neck/pelvis 1: Fracture or ligament injury in shoulder/arm/hand

4. Conclusion and discussion

Three hypothesizes was stated in the introduction; the following discussion will be based on the results and their implications for further research. Finally some reflections on the methodology and its' impact on the results will be addressed.

4.1 Self-reporters understand the phrasing of the questionnaire and can locate their accident on Google Maps

That 609 (590+19) respondents chose to locate the scene of their accident on Google Maps and successfully copy-paste coordinates into the questionnaire shows two important facts. First, that the task of using a computer for this (opening a new tab, pointing on map, copy-pasting) is feasible for a large proportion of the respondents. The input of a wrong GPS-format error (19 respondents of 833) could be considered as a methodological error and could have been avoided by making sure that the input field only recognized values entered in the correct format; it cannot be attributed solely to the respondents' lacking skills. Secondly, it shows that many of the respondents are willing to go to some lengths to share knowledge of their accidents; it must be emphasized that copy-pasting coordinates from Google Maps was voluntary in the questionnaire. That so many respondents chose to do this could be interpreted as an outcome of highly motivated respondents and their wanting to contribute to a safety research project – it is not given that the high level of respondents willing to locate their accident on Google Maps could be found in other studies of self-reported accidents. That the participants in this project were perhaps extraordinarily dedicated can be suspected from the high return rate – 97.5% of the monthly questionnaire about accidents was answered by

the participants (Lahrmann, Madsen et al. 2014). In comparison, a Swedish study on accidents with a questionnaire approximately the same length and with the same general themes as this, had a return rate of 35% (Tivesten, Jonsson et al. 2012). The Swedish study did not send out questionnaires each month but only once; seen in that perspective the differences in return rate are even higher.

Further studies are needed to conclude if the location pointed out by the respondent was also the correct location where the accident took place. This could be studied by for instance interviewing the respondents while focusing on determining the location, showing them pictures of the site they have selected and asking if it was the location of their accident or perhaps by accompanying them to revisit the accident location.

That only 6% of the respondent did not give accident descriptions consistent with their multiple choice answers of number and type of counterparts seems quite low. This indicates that the wording in the questionnaire seems rather easily understandable by respondents, which in part could be due to the explanatory text in the latter part of the question: "*a part could be a car, a truck, a bicyclist, a pedestrian, an animal and alike".* Still it must be kept in mind that not all respondents gave congruent answers; improvements could be made to further enhance their comprehension of the multiple choice question. However, some uncertainties arise when trying to establish the comprehension of the phrasing. The number of obvious misunderstandings can only be used as an indicator of comprehension, since it is also possible that the respondents misunderstood the word, but none the less chose the number of parties that matched the accident description. For instance, a car with only one driver could be viewed as one counterpart whether you think of the term *party* defined by the number of traffic elements or defined by the number of persons involved. Interviews with non-professionals could be used to further investigate the level of comprehension.

4.2 Self-reporters who claims to have been in contact with either police or hospital personnel in connection with their accident can be located in police or hospital records

Only 6 respondents could be located in police records, while 23 respondents answered that they thought a police report had been filed. The large discrepancy between expected records and actual records cannot be explained by findings in this study. Two explanatory reasons presents themselves; either the respondents did not have any contact with the police, but have another reason for believing that there would have been filed a police report (perhaps thinking polices reporting is connected with hospital admittance or insurance claims, or that the counterpart contacted the police) or the police was in fact contacted by the respondent, but the police chose not to file a report. It the latter is the case, erroneous reporting practice would be present by police personnel since the police must file a report when a person suffer injuries due to an accident. As 14 of the respondents, who thought a police report was filed when it was not, also sought medical attention – indicating that they had sustained some injuries from the crash – the discrepancy is quite concerning. There cannot be drawn any conclusions on the reason for the discrepancies; further studies would be needed. For instance by gaining access to the police recordings of all inbound calls to see if respondents made any contact that could lead them to suspect a report was filed.

99 respondents answer that they had medical care at the hospital or emergency room due to their accidents. 74 can be matched by their CPR-number with an accident (that could not be excluded as a possible traffic accident) that happened in the timespan of one month prior to the self-reported accident date and two months after. The discrepancy of 25 self-reported visits to the hospital and the number of visits noted in hospital records cannot be explained. Three explanations seem possible. First, the respondents could give

false information, in order to make their trips on bike seem more dangerous. The respondents were aware that the results would be used in a research project on bicyclist safety; social desirability effects could influence the respondents. The second explanation for the discrepancy between self-reported medical attention and hospital data could be due to recording errors at the hospital, as records were excluded from matchmaking if the recorded injuries and activity could not be considered a match for the self-reported accident. But it must be presumed that the medical personnel recording injuries at the hospital are rather precise when it comes to medical issues; an erroneous recording of injuries seems much more unlikely than an erroneous recording of e.g. the accident location. Thus recording errors could be expected to explain very few of the discrepant cases. Finally, errors in the respondents CPR-numbers would also prohibit matching. This could arise from typing errors by the respondents self-reporting their CPR-number or respondents willingly typing a false number if they did not want the researchers to know their personal identifier. That respondents reported more accidents than was to be found in the official datasets coincide with other research on self-reports and police recorded accidents (Arthur, Tubre et al. 2001), (Boufous, Ivers et al. 2010), but there are no studies on self-reported medical attention to compare the findings with.

The significant difference found between the average ratio of "police to hospital records" in Denmark and the ratio found in this study could indicate one of two things. Either the sample has more police reports than could be expected; for instance if the accidents were more severe or the respondents were more likely to contact the police than on average. Or it could be possible that the sample have more mislabels by the hospital personnel than would be expected. There are no other studies that could be used to compare the level of mislabels by hospital personnel, and it is thus not possible to make any conclusions on the reason for the statistical difference.

4.3 The self-reported accident information is congruent with information in the records written by hospital and or police personnel.

The 6 matching police records and the self-reported accident information seem fairly congruent, but due to the low number of matching records some reservations persist.

The hospital records on the other hand is not very congruent with the self-reports. Of the 74 matches, only 53 were registered with a traffic accident in LPR. As it is the accidents recorded as traffic accidents in LPR that are used for statistics on the level of underreporting (Danmarks Statistik 2014), this finding indicates that the level of underreporting by the police could actually be much higher than previously thought due to the mislabeled reports in the hospital database. A possibility exists that the amount of injuries not labeled as traffic injuries to some degree could be counterbalanced by people seeking medical care with injuries sustained by something other than a traffic accident being mislabeled as victims of traffic accidents. This is mentioned as a source of error by (Roberts, Vingilis et al. 2008); the E-codes of a person seeking medical attention after a traffic accident, but with injuries not originating from the accident would be at risk of being registered with the injuries as accident-related by mistake. That hospital data seems to suffer from underreporting as well as police records is something not previously researched in Denmark and should be subjected to further research.

4.4 Methodological issues and self-reports

Some concerns regarding self-reporting in traffic research were mentioned in the introduction; disagreement with other data sources, social desirability effects, statistical bias and forgetting.

The findings in this study shows that at least some of the disagreement between other data sources and self-reports cannot be attributed to low validity of self-reports. This is evident in the discrepancy between the 74 matches in LPR and the 53 recorded as traffic accidents in LPR. As the 21 records not recorded as traffic accidents are obviously mislabeled, one cannot conclude that low disagreement between hospital records and self-reported accident information could be attributed to erroneous self-reports.

As the original research project had to do with bicycles, the respondents are all bicyclists. In some countries, where the use of bicycles aren't that common, a person who chooses to transport him or herself by bicycle could be suspected to have other personal traits than the average car user (eg. more environmentally aware, a feeling of being one of a kind or perhaps identifies strongly with being a bike rider). As bicycles are a normal mode of transport in Denmark for both adults and children, one could argue that this difference is not that pronounced in the current study. But none the less, the participants in this study is bicyclists who volunteered to drive with a brightly colored jacket for a year – they are suspected to be very interested in traffic safety and highly motivated to contribute to the project's studies. The near-misses reported as accidents could be seen as an indication of a form of social desirability effect where respondents wants to give the socially most desirable answer (Wåhlberg, Dorn et al. 2010) – in this case thinking that researchers have the notion that traffic is dangerous and thus conforming their answers to this point of view. As making up accidents would be outright lying, it would not be considered desirable behavior thus leaving the respondents to report near-misses in order to be "thorough respondents". This could be the reason for the 58 of 833 self-reported accidents actually classifies as near-misses. But as no other studies on self-reported accidents and near-misses have been found it is not possible to conclude whether the respondents report more or less near-misses than expected. The social desirability effects associated with the respondents' voluntarily participation in the research project could as mentioned before also affect the level of claimed medical care. Also the statistical significant difference of the ratio of police-to-hospital records could be affected by the biased group of respondents; perhaps they are more concerned with safety and thus more likely to contact the police than the average bicyclist.

The concerns regarding possible over-reporting by some subgroups as mentioned by (Wåhlberg 2009) cannot be addressed properly with the available data; it cannot be rejected that the volunteers could be considered an "over-reporting subgroup". This would to some degree explain the discrepancies between official records and the self-reported contact to hospital or police. It cannot be rejected, but it seems in contradiction with the concept of social desirability that the participants would have been making up false accidents and thus over-reporting.

Memory is an important factor in the validity of the given self-reports. The recall period in this study is a month or less due to the questionnaire being sent out on a monthly basis. There is no agreement on how much recall period affects the validity; some sources claim that more than two months is expected to affect the validity and reliability (Roberts, Vingilis et al. 2008), and some find no significant differences of correctness of accident recall depending on recall delay (Versteegh 2004). As most studies tend to compare official data with self-reported number of accidents with a recall between one or five years, there is no literature that can support the hypothesis that a recall period of only a month should not contribute to respondents simply forgetting an accident and thus not reporting is. But as the respondents are asked every month, telescoping effects (remembering an accident as more recent than it actually is) can only affect the first month of answers.

Literature

ARTHUR, W.J., TUBRE, T., DAY, E.A., SHEEHAN, M.K., SANCHEZ-KU, M.L., PAUL, D., PAULUS, L. and ARCHULETA, K., 2001. Motor Vehicle Crash Involvement and Moving Violations: Convergence of Self-Report and Archival Data. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* **43**(1), pp. 1-11.

AUSTIN, K.P., 1994. *The Identification of Mistakes in Road Accident Records – Part Two.* 407. Institute of Transport Studies, University of Leeds.: White Rose University Consortium.

AUSTIN, K.P., 1993. *
The Identification of Mistakes in Road Accident Records – Part One: The Use of Geographic Information Systems.* Working Paper 406. Univerity of Leeds, Institute of Transport: White Rose University Consortium.

BOUFOUS, S., IVERS, R., SENSERRICK, T., STEVENSON, M., NORTON, R. and WILLIAMSON, A., 2010. Accuracy of self-report of on-road crashes and traffic offences in a cohort of young drivers: the DRIVE study. *Injury Prevention,* (16), pp. 275-277.

BROUGHTON, J., KEIGAN, M., YANNIS, G., EVGENIKOS, P., CHAZIRIS, A., PAPADIMITRIOU, E., BOS, N.M., HOEGLINGER, S., PÉREZ, K., AMOROS, E., HOLLÓ, P. and TECL, J., 2010. Estimation of the real number of road casualties in Europe. *Safety Science,* **48**, pp. 365-371.

BULL, J.P. and ROBERTS, B.J., 1973. Road Accident Statistics - A Comparison of Police and Hospital Information. *Accident Analysis and Prevention,* **5**, pp. 45-53.

DANMARKS STATISTIK, 2014-last update, Tabel MOERKE. Available: http://www.statistikbanken.dk/MOERKE [28.Maj, 2015].

ELVIK, R. and MYSEN, A.B., 1999. Incomplete Accident Reporting - Meta-Analysis of Studies Made in 13 Countries. *Transportation Research Record,* (1665), pp. 133-140.

FINESTONE, H.M., GUO, M., O'HARA, P., GREENE-FINESTONE, L., MARSHALL, S.C., HUNT, L., JESSUP, A. and BIGGS, J., 2011. Department of Transportation vs Self-reported Data on Motor Vehicle Collisions and Driving Convictions for Stroke Survivors: Do They Agree? *Traffic Injury Prevention,* **12**(4), pp. 332-337.

HEMDORFF, S., LUND, H.V. and TAUL, B., 2003. *Indberetning af færdselsuheld: Vejledning.* Raport 277. Copenhagen, Denmark: Vejdirektoratet.

ISAKSSON-HELLMAN, I., 2012. A Study of Bicycle and Passenger Car Collisions Based on Insurance Claims Data, *6th AAAM Annual Conference, Annals of Advances in Automotive Medicine, October 14-17, 2012* 2012.

JANSTRUP, K., HELS, T., KAPLAN, S., SOMMER, H. and LAURITSEN, J., 2014. Understanding traffic crash under-reporting: linking police and medical records to individual and crash characteristics. *Transport Research Arena 2014, Paris,* .

LAHRMANN, H., MADSEN, J.C.O., MADSEN, T.K.O., OLESEN, A.V., HANSEN, S., THEDCHANAMOORTHY, S. and BLOCH, A., 2014. *Projekt Cykeljakken: Den sikkerhedsmæssige effekt af en gul cykeljakke.* DCE Technical Report nr. 167. Aalborg: Aalborg University, Department of Civil Engineering.

LAHRMANN, H., MADSEN, T.K.O., OLESEN, A.V. and MADSEN, J.C.O., 2014 - submitted to journal. The safety impact of a yellow bicycle jacket. *Accident Analysis and Prevention,* .

LAJUNEN, T. and ÖZKAN, T., 2011. Self-Report Instruments and Methods. *Handbook of Traffic Psychology.* Elsevier Inc., pp. 43-59.

ROBERTS, S.E., VINGILIS, E., WILK, P. and SEELAY, J., 2008. A comparison of self-reported motor vehicle collision injuries compared with official collision data: An analysis of age and sex trends using the Canadian National Population Health Survey and Transport Canada data. *Accident Analysis and Prevention,* **40**, pp. 559-566.

SUNDHEDSSTYRELSEN, 2009. *Fællesindhold for basisregistrering af sygehuspatienter 2010.* www.sst.dk: Sundhedsstyrelsen.

TIVESTEN, E., JONSSON, S., JAKOBSSON, L. and NORIN, H., 2012. Nonresponse analysis and adjustment in a mail survey on car accidents. *Accident Analysis and Prevention,* **48**, pp. 401-145.

TIVESTEN, E. and WIBERG, H., 2013. What can the drivers' own description from combined sources provide in an analysis of driver distraction and low vigilance in accident situations? *Accident Analysis and Prevention,* (52), pp. 51-63.

VERSTEEGH, S., 2004. The Accuracy Of Driver Accounts Of Vehicle Accidents, *2004 Road Safety Research, Policing and Education Conference. Session: Crash and Injury Analysis.* 2004.

WÅHLBERG, A.E.A., 2010. Social desirability effects in driver behavior inventories. *Journal of Safety Research,* **41**, pp. 99-106.

WÅHLBERG, A.E.A., DORN, L. and KLINE, T., 2010. The effect of social desirability on self reported and recorded road traffic accidents. *Transportation Research Part F: Traffic Psychology and Behaviour,* **13**(2), pp. 106-114.

WÅHLBERG, A.A., 2009. *Driver behaviour and accident research methodology: unresolved problems. (Human factors in road and rail transport)* . Farnham, England: Ashgate.