

# **PROPOSAL FOR FLEXIBLE HARMONIZED TRUCK IN-DEPTH ACCIDENT DATA COLLECTION METHOD WITH EXPOSURE CONTROL GROUP FOR ACCIDENT CAUSATION.**

**Y.W.R. de Vries, N.A. Jongerius**

TNO Automotive, Delft, The Netherlands  
DAF Trucks, Eindhoven, The Netherlands

## **Presenter**

Y.W.R. de Vries M.Sc.  
Schoemakerstraat 97,  
P.O.Box 6033,  
2600 JA DELFT  
The Netherlands  
+31 (0)15 269 73 09  
ydo.devries@tno.nl

## **ABSTRACT**

**Internationally, the need is expressed for harmonized traffic accident data collection (PSN, PENDANT, etc.). Together with this effort of harmonization, traffic accident investigation moves more and more in the direction of accident causation. As current methods only partly address these needs, a new method was set up. The main characteristics of this method:**

- All relevant accident data can be included: human-, vehicle-, and environmental related data for the pre-crash, crash and post-crash situation (the so called Haddon matrix). The level of detail can be chosen depending on interest and/or budget, which makes the method very flexible.**
- Accident / injury causation (associated) factors can objectively be identified and quantified, by comparison with exposure information from a control group population.**
- International harmonization can be achieved, because the database fields are maximally matching the best elements of the existing databases.**

**In this paper the accident collection and control group method are presented, including some of the achieved results from a pilot study on 30 truck accidents and 30 control locations. The data were analyzed by using cross-tabulations and classification-tree analysis. The method proved useful for the identification of statistically significant causal aspects.**

# PROPOSAL FOR FLEXIBLE HARMONIZED TRUCK IN-DEPTH ACCIDENT COLLECTION METHOD WITH EXPOSURE CONTROL GROUP FOR ACCIDENT CAUSATION.

Y.W.R. de Vries, N.A. Jongerius

TNO Automotive, Delft, The Netherlands  
DAF Trucks, Eindhoven, The Netherlands

## 1 INTRODUCTION

Internationally, the need is expressed for harmonized traffic accident data collection (PSN, PENDANT, etc.). Together with this effort of harmonization, traffic accident investigation is used more and more to address all cells of the Haddon-matrix (see Figure 1). The information is used to deploy new activities in relevant areas.

Currently available databases are generally not sufficiently similar to allow an efficient and effective exchange/combination of accident data from different sources. Furthermore, many of these databases find their roots in secondary safety, so their structure is built around accident consequences with some extensions towards tertiary safety.

Knowledge on primary safety aspects (i.e. avoiding accidents) requires information about accident causation. Databases and methods that have been developed for accident causation studies (e.g. EACS, ETAC), belong to the case-series studies: cases are investigated and frequency counts can give information on occurrence of possible risk factors in these accidents. The impossibility to relate these occurrences to reference data is a large drawback as will be shown later.

A new method has been set up, with the main aim to address the aforementioned limitations of current databases. The main goals of this paper are to discuss the principles of this newly developed method for an epidemiological study into accident causation, and to show the results of a first pilot analysis on truck accidents.

The work has been carried out by TNO, under support from DAF Trucks, Scania Trucks, and the Dutch Ministry of Transport, Public Works and Water Management and the Ministry of Economic Affairs.

### 1.1 Exposure

Correct information about the cause(s) of accidents is relevant for policy makers and vehicle manufacturers. Identification of accident causes requires the acquisition of large amounts of data per accident, and a correct interpretation of this data (accident reconstruction, etc.).

	<i>Human</i>	<i>Vehicle</i>	<i>Environment</i>
<i>Pre-Crash</i>	Prevention	Crash avoidance	Road Infrastructure maintenance/design
<i>Crash</i>	Biomechanics	Crashworthiness	Crashworthiness
<i>Post-Crash</i>	Acute care and rehabilitation	Prevent fires automated collision notification	Emergency medical services

Figure 1. Example of the Haddon matrix

Incorrect knowledge on accident causes may lead to the implementation of non-effective countermeasures. Such incorrect knowledge may be due to exposure effects. A short example can illustrate why an effort to simply reduce the highest frequency of occurrence could stumble upon some problems.

*Example:*

*Suppose one wants to reduce the relative high frequency of right turning truck vs. bicycle accidents, which occur in urban areas (on average 17 fatalities and 44 seriously injured per year in the period 1997-2000 in the Netherlands, [DOT,2005]). Can it be expected that a blind-spot mirror will reduce the number of accidents? That depends on whether the accident probability can really be influenced by blind-spot mirrors. It could be the case that drivers normally anticipate very well, but that under accident circumstances attention diverting or unexpected environmental aspects are present, which are not or hardly present under normal conditions. If this is the case, blind-spot mirrors can not be expected to have a large influence, while solving these other aspects should have effect.*

So, some kind of reference data or control population is needed to correctly identify risk factors. A literature survey was carried out to find previously used study setups which make use of reference data. Possibilities for obtaining such observational data are the following:

Internal control groups:

Internal control groups are groups of accidents for which a specific parameter is assumed not to have any influence. Differences in the presence or absence of the parameter in the studied group and the control group can indicate a relationship with accident occurrence. Two main problems exist: many cases have to be present and it has to be sure that the parameter under study has no effect in the control group.

Global indicators:

Sometimes global indicators (kilometers driven, etc.) are used as reference data to indicate potential problems. However, the results are very dependent on the indicator that is used [Elsevier, 1997] and can be tuned with the use of an indicator which provides the results that are wanted. Furthermore, global indicators can not go into the detail needed for accident causation research (e.g. type of use).

Cohort study:

Because traffic accidents have a relatively low occurrence, cohort studies are inefficient. For accident research a group of drivers should be followed for a certain time. During this time some accidents should occur in this group. Drivers with and without accidents can be compared. The presence of accidents in this group is expected to be quite low when the group is not very large or the study duration is not very long.

Case-control study:

When a case-series study is extended with the collection of some sort of control group, which can be used as reference data a case-control study design is obtained, from which associations between factors can be obtained.

Because in-depth research is already a case-series study, extension to a case-control study is therefore the most logical approach.

Why hasn't this been done before? Some attempts and good starts have been made:

- A recent example is the European Motorcycle Accident In-Depth Study (MAIDS) [OECD, 1999], carried out in five countries. However, this study only looked at vehicle technical aspects, driver, clothing and behavior, and did not take into account environmental aspects such as road type, road widths, traffic density, etc. By using a petrol station inquiry method, possibly not the right exposure was obtained because of different gas-tank sizes and fuel consumption rates. Secondly driver co-operation is an issue.
- Another option sometimes used is to question drivers passing through the same scene one week after the accident. In this case, also the same environment is studied as in the accident situation. Analysis on environmental factors is therewith no longer possible. Secondly, it is hardly possible to get co-operation of drivers.

However, both of these methods provide a brilliant start for something new.

## 2 METHODS

### 2.1 Interaction model

What we would like is to obtain an exposure measure containing vehicle, human and environmental factors. These factors interact in an unknown way, and therefore should be investigated at the same time (see also Figure 2).

The vehicle – environment relationship indicates that there is a possible interaction between the two. Kippers are likely to drive in other areas than long-distance transport vehicles.

Vehicle and driver also have possible interactions. Behavior may be different with full load or empty vehicle, or even with dangerous goods.

Driver – environment interaction is possibly the most complicated. This means that some sort of driver behavior or physical status of the driver is related to a certain type of environment (e.g. using cruise control on motorways only, driver irritation due to traffic circumstances).

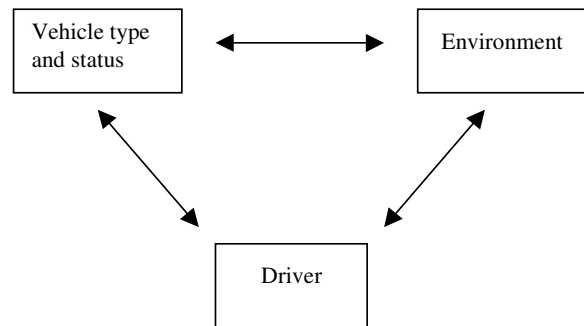


Figure 2. Main interactions.

### 2.2 Investigation approach

In theory one would like to obtain a control group population from a normal traffic situation, which can directly be compared with the accidents: some kind of “virtual accident”. Traffic intensity can be used as a measure. The number of “virtual accidents” is then calculated by:

$$N = n_m((n_m-1) + n_o + \text{one environmental impact}) \quad (1)$$

In which  $n_m$  equals the number of vehicles in the main direction and  $n_o$  the number of vehicles in the other possible directions.

The traffic counts can be obtained from e.g. video, together with driven speeds, maneuvers, distances to other road users, color, etc. From all the vehicles passing through the location the drivers should be observed and interviewed, and the vehicles should be inspected. The traffic locations that would be used should be a random sample of the roads available in the

region. The total of all virtual impacts in all monitored traffic situations will then serve as the control group population under exposure.

This method of gathering all data at one time is unfortunately practically impossible, because not all vehicles can be stopped in a monitored scene. However, there is a way around this when some conditions and assumptions are met.

### 2.3 Model assumptions

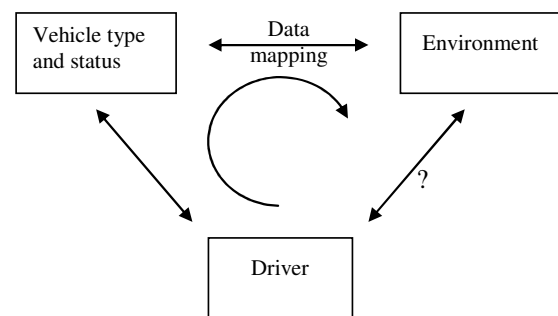
The environmental conditions and exposed vehicles can be observed with video at randomly selected locations in the area in which the accidents were collected. From this information “virtual accidents” can be generated. This process will be described later.

The main problem at this point is that all data still can not be obtained at one time. Two possibilities now arise. For (semi) permanent physical conditions, for which we expect that they have no relation with the environmental conditions (e.g. gender, illness, etc.) a control group population could be gathered at any given location. Transient conditions which may have an interaction with or are induced by the environment are most complicated (attention diversion, using cruise control on motorways). By interviewing people about the frequency of use or habits non-environment related issues can be investigated. Environmentally related habits can also be investigated on a more global level by asking about frequency of use under specific circumstances. This method suffers (slightly less) under the same problems as global indicators.

The other option could be to form a cohort of random drivers whose behavior is recorded / logged under occurring circumstances. This option is more complicated and time consuming, but the result is likely better. For financial reasons the interviewing method is chosen for this study.

The driver information will be obtained at convenient locations which are sufficiently randomized. Drivers are interviewed and vehicles inspected.

The idea that is now used is the following (see Figure 3): The driver interviews and detailed vehicle inspections are treated as missing values in the data from the video observations. These missing data are imputed from the separate vehicle inspections and interviews on the video information. The random imputation is conditioned (matched) on variables collected in both inspections (so called conditional random imputation). These values may not be treated as real values per accident but can be used to make appropriate inferences and to show statistical associations. The vehicle type distribution does not have to match the one found from the video observations and can deliberately be biased towards groups of interest to obtain the most useful information. This results in more detailed information for the vehicles of interest. The parameter for which the bias is introduced should also be measured from the video data (vehicle type, color, etc.). It would not lead to a biased sample, because the distribution of generated virtual accidents would stay the same, only with more or less details for specific groups.



**Figure 3. Diagram of the interactions between driver, vehicle and environment.**

In future projects multiple imputation techniques can be used in order to improve the prediction of the missing values [Rubin, 1987]. For this pilot study this method has not been used yet.

Isn't imputation just making up data? It is when the values are treated as real values, but used in statistical analysis it gives a correct prediction for distributions that can be used for comparison with other distributions.

## 2.4 Analysis

The analysis focuses on under and over represented factors in the comparison of accidents and virtual accidents. This requires that no biases should exist in accident notification and collection, otherwise representation differences are due to poor sampling.

For the analysis SPSS was used.

## 2.5 Practical implementation

### 2.5.1 Location selection

The inspection of the locations should be completely randomized over time and the research area. Depending on the research question, several biases could be introduced to exclude known or not-interesting information. For the whole of the Netherlands information on all roads is available from a Geographic Information System (GIS) database (see Figure 4). Most other countries also have mapped their (main) roads into vector based files that are compatible with GIS [GIS, 2004]. The location selection has been split into crossings and segments without crossings. In order to sample all roads equally, all segments have been divided into 25 meter sections. Sampling the number of lanes instead of number of roads is more appropriate, so that the traffic flow on one-way streets and multiple-lane streets are sampled just as much as on two-way streets.



**Figure 4.** Example of a GIS-location and the low visibility video pole clamped to a lamp-post.

For crossings the following approach is taken. Each lane into the crossing is counted as an intersection, because vehicles can come from all directions. Roundabouts are considered to be set of T-crossings with two in-coming lanes and one out-going lane. This is the same as considering it to be one crossing, due to the fact that the number of maneuvers is limited in the multiple T-crossing approach.

### 2.5.2 Obtaining the information

In Figure 5 the followed workflow is depicted. The selected lane will be the main traffic stream. All lanes and the scene are recorded on video for approximately 30 minutes from a high location (an extendable beam) to reduce parallax effects in the analysis. With special developed software and markings on the road with known distance to each other, the speed and distance to other vehicles can be obtained from the video. For trucks extra information is recorded that will improve the conditioning (matching) for the required conditional random imputation (see 2.3).

The frequency of passing vehicles is considered the control group population, which all potentially can have an accident.

The interviews were done at restaurants, gas-stations, distribution centers and companies using (specific) trucks. The selection of truck types was matched to the distribution observed in the accident sample, again to obtain a maximum of detail with minimum effort. During the analysis a comparison will have to be made between accident trucks of the same type (e.g. on mirror adjustments). Not all truck types could be investigated with the limited sample.

## 2.6 Virtual accidents

In order to compare accidents with exposure information it has to be in the same format. Therefore, virtual accidents are created from this information. Each target vehicle (in this case a truck) has a virtual accident with every other vehicle passing through the scene. The number of virtual impacts is therewith frequency induced. The more other vehicles are present, the more virtual impacts. Environmental information was coded with the virtual accident.

Each vehicle can have an impact with a vehicle in the same direction as well as with a vehicle from another direction or with an environmental object.

A main problem with generating virtual impacts is that the number increases quadratic with each extra vehicle in the main stream. Therefore the number of generated virtual impacts is limited to ten thousand per location.

## 2.7 Data weighting

Because a sample is now taken from what really occurred in these 30 minutes of video, the data has to be weighted accordingly. Another problem which manifested itself is that due to several reasons the video duration of 30 minutes was in more than one occasion not 30 minutes. Also analyzing 30 minutes of motorway video at this time is undoable, and therefore a sample had to be taken. This made is necessary to correct the obtained number of virtual impacts.

After establishing these weight factors per impact, the locations still had to be weighted towards their real presence in the sampling region.

## 2.8 Database

For the purpose of this study a new database was developed (see Figure 5). The reasons to do this are that there exist many different databases which are incompatible with each other, while one of the aims of this study was to create an internationally harmonized study setup. The database also had to be able to incorporate the information gathered from the control group population

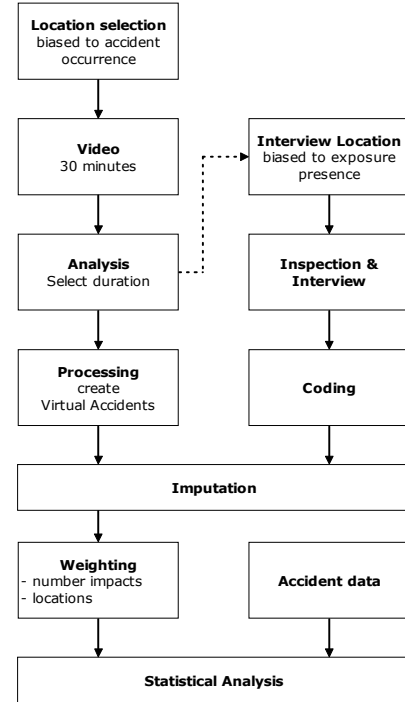


Figure 5. Work flow diagram.

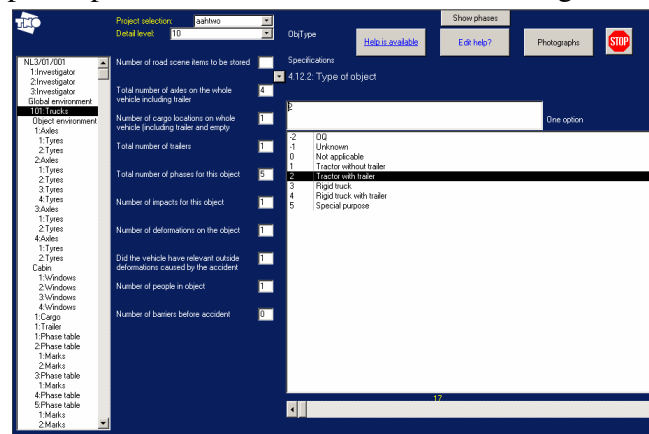


Figure 6. Screen shot of the database.

(the virtual accidents). The developed database incorporates the main questions and answer options of methods used over Europe: Pendant, EACS, OECD-Common Methodology for motorcycle accidents, Gidas. Obviously some choices had to be made for the best questions. The database is flexible: depending on the available finances some level of detail can be chosen and new questions and options can be added. Specific questions for trucks have been added. Multiple impacts can be coded, impact partners are linked, and contact and deformation codes are present. Each item that can be present multiple times (vehicles, tires, victims, injuries, etc.) can be coded unlimited. The database is free of use, and can be obtained via the authors.

### 3 RESULTS

Because the sample of accidents is limited, the authors want to stress that the results presented here can only serve as an indication and merely show the analysis possibilities.

#### 3.1 Accidents and control group population

For this study 30 truck accidents were investigated. The Dutch Accident Research Team (DART) was notified by the technical police departments in the province of Zuid-Holland. The technical police departments in these regions are notified of all truck accidents and measure accident (skid) marks in detail. All accidents for which TNO was notified were investigated by the investigation team. The total sample is biased towards more severe accidents. Comparison with the control group therefore gives information about severe accidents.

Of these accidents, 15 occurred on straight segments and 15 on crossings. The locations of the control group population were matched with the accident

occurrence locations and weighted to obtain a maximum of accuracy and detail.

#### 3.2 Encountered problems

- GIS – information is sometimes incorrect (selection of bicycle or foot paths), roads are under construction or it is impossible to film at a given location. A backup list of locations is therefore needed.
- During interviews it sometimes takes a long time before a truck arrives. This is unavoidable because it is a Poisson process and should be taken into account
- Video data acquisition at night causes some problems which should be solved.

**Table 1. Comparison between collision partners in accident cases and in virtual accidents on intersections (The distributions differ significantly. Chi<sup>2</sup> test: p<0.05). Cases represent accidents and controls represent virtual accidents.**

			Case		Total
			Case	Control	
OV1: Object type	Truck	Count	0	14158	14158
		% within Case	.0%	6.6%	6.6%
		Adjusted Residual	-1.0	1.0	
Car	Count	3	170786	170789	
	% within Case	20.0%	79.6%	79.6%	
	Adjusted Residual	-5.7	5.7		
Motorised 2-wheeler	Count	3	354	357	
	% within Case	20.0%	.2%	.2%	
	Adjusted Residual	18.8	-18.8		
Bicycle	Count	6	1008	1014	
	% within Case	40.0%	.5%	.5%	
	Adjusted Residual	22.3	-22.3		
Bus	Count	1	3145	3146	
	% within Case	6.7%	1.5%	1.5%	
	Adjusted Residual	1.7	-1.7		
Van	Count	1	24900	24901	
	% within Case	6.7%	11.6%	11.6%	
	Adjusted Residual	-.6	.6		
Other vehicle	Count	0	8	8	
	% within Case	.0%	.0%	.0%	
	Adjusted Residual	.0	.0		
Pedestrian	Count	0	12	12	
	% within Case	.0%	.0%	.0%	
	Adjusted Residual	.0	.0		
Stationary object	Count	1	166	167	
	% within Case	6.7%	.1%	.1%	
	Adjusted Residual	9.2	-9.2		
Total	Count	15	214537	214552	
	% within Case	100.0%	100.0%	100.0%	

a. Junction = Yes

### 3.3 Example analysis

In Table 1 the frequencies of collision partners in the accident cases and in the virtual accidents (control) on intersections are shown. The adjusted residuals indicate whether the cases and controls differ significantly and may be interpreted as follows: an absolute value larger than two indicates for normal distributions a 95% certainty that a significant difference is present [SPSS, 1998]. The sample in this pilot-study is unfortunately too small to satisfy the condition of normality. From the presented table it can be read that motorized two-wheelers and bicycles are more present in the accident cases than in the control group. If the sample were large enough, this would indicate that the probability to be in an accident as a bicyclist or motorized two-wheeler is higher than for example a car-driver.

A classification tree analysis [SPSS, 1998] with a forced split on vulnerable road users (motorized two-wheelers, bicycles and pedestrians) was used to identify aspects in truck-vulnerable road user impacts, which were found to have a high probability of occurrence (see appendix). The variables in the classification tree were the maneuvers of the truck and vulnerable road user with respect to each other.

Compared with the virtual truck accidents (frequency of normal meetings) with vulnerable road users (VRU) two groups can be identified with clear differences in occurrence:

- Truck turning right, with the original driving direction identical to the VRU driving direction. This situation was never observed in the virtual accidents. Not shown is that this occurs in all investigated cases on local small roads and that the VRU is going straight or is turning right. This is typically the situation in which blind angle aspects are considered relevant.
- Truck is going in the opposite direction from where it comes from (driving backwards); while in the control group this situation is never observed. This is possibly also a blind angle aspect, but on the rear of the truck.

## 4 DISCUSSION

### 4.1 On the method

An aspect that is more difficult to investigate with this study setup is the environmentally related driver behavior: whether an association exists between certain driver behavior and the environment (e.g. of cruise control on motorways, use of mirrors at certain locations). When it is expected that these factors play an important role, the relationship can be investigated by implementing questions regarding these relationships in the interviews. Frequency of use under various conditions can be asked. Another more expensive method already mentioned, is to form a cohort of random drivers for which the behavior and actions are recorded in some way, possibly by actually monitoring the driver and recording and coding the behavior. Again it is not required that the population matches the exposure information from video.

Night time control samples caused some problems. Video information with “night shot”-mode was not of very good quality. Improvements have to be sought in infrared sensitive cameras.

### 4.2 On the results

From the classification tree in the appendix it can be seen that right turning trucks and VRU's coming from the same direction have a relatively high number of cases (3) with respect to the control group (0). At the top of the classification tree this was (10 cases/ 250 controls). However, if due to chance a location near a school was sampled in the control group or sampled at hours at which children and students bike to school, many more virtual bicycle accidents

would have been present. Therefore no real conclusions may be drawn from this sample. When the sample would have been large enough and the same situation would persist it might have been concluded for these cases that situations with right turning trucks and VRU's coming from the same direction impose a greater accident probability than other cases. It then could be suggested that this relates to blind angle aspects.

With more cases the classification tree analysis could go further. If any control group cases would be present for these typically dangerous situations a comparison between environmental, driver or truck related issues could be made for these locations to show typical problems. But at this time, no control data is available and the number of accidents is clearly limited.

Although potentially influenced by the coincidental choice of control locations, the method of analysis seems to indicate a potential risk factor that was also identified in national statistics [Vries, 2000]. The conclusions from the national statistics were based on assumptions about exposure. This in-depth analysis shows that with control-group information this can be supported objectively. Details concerning mirror adjustment, road layout can give more details about exact causation related aspects. Again comparison with control group information can show discrepancies between the two data sets. This information can be further supported by objective and subjective descriptive information.

## **5 CONCLUSIONS & RECOMMENDATIONS**

From the literature study it was learned that a case-control study is best suited for in-depth traffic accident research at this time. No good documented case-control study could be found, which includes environmental, driver and vehicle information. Therefore the new method was developed.

- Injury causation analysis can be done in great detail. A large amount of data analysis possibilities exist and seem to give good information and indications to find problems in accident and injury causation from which new solutions may be derived.
- Risk adaptation for primary and secondary safety features can be assessed.
- Environmental, human and vehicle factors can be investigated together, taking into account the relationship between the factors.

### **5.1 Recommendations**

When defining measures for improved safety, it is recommended to include a dedicated exposure evaluation in order to determine with statistical significance whether, and up to what extent, actual safety improvements can be expected. The case control method presented here is a good approach.

The exposure method should be evaluated on a wider scale, preferably European wide to effectively indicate risk factors. European projects like SafetyNet or TRACE might provide a good basis.

## 6 REFERENCES

- Bangdiwala, S.I. (2000). "Methodological Considerations in the Analysis of Injury Data: A challenge for the Injury Research Community", Injury Prevention and Control, Taulor & Francis, ISBN 0-748-40959-9.
- DOT (2005). "<http://www.minvenw.nl/dgg/dodehoek/Zichtveldprobleem/Feitenencijfers/#feiten>"
- Elsevier (1997). "De kans op een ongeluk", 6-9-1997
- GIS (2004). "<http://www.gis.com/whatisgis/index.html>"
- OECD (2001). "Motorcycles: Common international methodology for in-depth accident investigations"
- Rubin, D.B. (1987). "Multiple imputation for non-response in surveys.", John Wiley & Sons, New York.
- SPSS (1998). "AnswerTree 2.0 user's Guide", ISBN 1-56827-254-5
- Vries, Y.W.R de (2000). "Accident analysis of heavy trucks", TNO Report, 00.OR.BV.053.1/YdV
- Vries, Y.W.R de (2005). "A method for control group data to find accident- and injury causation factors in in-depth traffic accident studies", TNO Report, 05.OR.SA.034.1/YdV

# 7 APPENDIX

Read underneath vulnerable road users: From the Other Vehicle (OV) perspective the truck is going into a certain direction relative to the OV. Suppose the truck is turning right. Then read: From the OV perspective the truck is coming from, e.g. the identical direction as the OV.

